COK-H05003-50-P009644

ISSN 0001 253X
Volume 52
Number 1
January 2000

50

1.7.2000

Proceedings

New information perspectives

Comment
Changing names
Harry East

Tales on the Internet: making it up as you go along
Chris Frost

A strategic approach to research using Internet tools and resources
Lyn Robinson

Task analysis of the webmaster: results of an empirical study
Pieter Van der Walt & Pieter Van Brakel

Teaching online information retrieval to students of journalism
Charlotte Wien

Aslib
THE ASSOCIATION FOR INFORMATION MANAGEMENT

Aslib Proceedings is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) are very much in mind. Articles are refereed to meet the need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In exceptional cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000–6,000 words in length. Articles should be accompanied by a 100–150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The editor is happy to provide advice on an informal basis to potential authors.

Editorial Board

P9644

Dr David Nicholas, *Head of Department, Department of Information Science, City University*; John Akeroyd, Director, *Learning Resources, South Bank University*; Sarah Blair, *Head of Publications, Aslib*; Peter Chapman, *Information Manager, NewQuest North East*; Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire*; Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw*; Harry East, *Senior Research Fellow, Department of Information Science, City University*; Mike Hyman, *Managing Director, Head Software International*; Robin Hunt, *Creative Director, AreHause*; Nat Lievesley, *Network Manager, Centre for Policy on Ageing*; Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Lecturer, Department of Information Science, City University*; Richard Withey, *Director, New Media, News International*; Ralph Adam, *Freelance Journalist*.

Editorial Team

Editor: David Nicholas
Deputy Editor: Ian Rowlands
Editorial Manager: Kay Nicholas
Editorial Assistants: Peter Williams (City University), Colette Coles (City University)

010·5
AS 52 Proc

CONTENTS

3–4 Comment
Changing names
Harry East

5–10 Tales on the Internet: making it up as you go along
Chris Frost

11–19 A strategic approach to research using Internet tools and resources
Lyn Robinson

20–38 Task analysis of the webmaster: results of an empirical study
Pieter Van der Walt and Pieter Van Brakel

39–47 Teaching online information retrieval to students of journalism
Charlotte Wien

Comment
Changing names

Harry East
Department of Information Science, City University, Northampton Square, London EC1V OHB
Harry@soi.city.ac.uk

> 'It's a stupid name enough!' Humpty Dumpty interrupted impatiently. 'What does it mean?'
> '*Must* a name mean something?' Alice asked doubtfully.
> 'Of course it must,' Humpty Dumpty said with a short laugh: '*my* name means the shape I am . . . With a name like yours, you might be any shape, almost.'
> **Through the Looking Glass *and what Alice found there***

In a recent issue of *Bulletin of the American Society for Information Science*, its President-elect (Eugene Garfield, no less) calls for a 'symbolic step accompanied by concrete actions' starting with a name change of the Society from ASIS to ASIST. The T is for Technology [1]. He suggests that, apart from it being more descriptive of what many of its members do, the modification might also revive a flagging membership. Garfield recalls that remodelling the original name of the Society – the American Documentation Institution (ADI) – involved 'discussions that in those days were rancorous. The word *documentation* had lost its cachet. The then seemingly amorphous and controversial term *information science* was adopted'.

Rancorous discussions focussing on names and images are not unknown on this side of the Atlantic. For example, it seems that after a great deal of provocative foreplay, a majority of the Institute of Information Scientists' members voted at a recent AGM to get into bed together with the Library Association [2]. It remains to be seen that, if and when they do, what name they might give to the issue with which their union has been blessed. As in the United States, most of the British schools of librarianship in academia have gradually embraced the term *information* in their names. I suspect this was not achieved without internal strife. Some years ago Aslib tacked on the supernumerary 'Association for Information Management' to its well-established name, but it is still not used much.

This editorial is mainly about names. And it is fair to point out initially that the institution under discussion here – one that has traditionally funded library and information research – has undergone name modifications mainly as a result of changes in the parent bodies sponsoring it. In the early Sixties, the Department of Scientific and Industrial Research (DSIR) was the main public body supporting research in the field. The following are its lineal descendants.

Office of Science and Technical Information	(OSTI, 1965)
British Library Research and Development Department	(BLRDD, 1974)
British Library Research and Innovation Centre	(BLRIC, 1996)
BLRIC merged with the Library and Information Commission	(LIC, 1999)
And a recent announcement informs us that the research function is transferring to Museums, Library and Archives Council	(MLAC, 2000)

Discerning observers will note that, in spite of the tendency to aggrandisement of title,

Office → Department → Centre → Commission → Council

there has been a general decrease in the longevity of the listed bodies. Perhaps this reflects little more than the rapid change affecting many institutions. Recent alarms and excursions more certainly signify a new government with a revised agenda, aimed at a broader demographic and democratic community. However, the absence of the terms *information* or *research* from the title of the Council – one or the other in evidence in precursor organisations – 'must [like Hamlet] give us pause'. Their displacement by the museums-library-archives trinity suggests a reassertion of supposed eternal verities.

The advent of the MLAC seems, at first sight, a regression to a 'repository' concept, albeit in electronic and interactive form. What will be the future of research in information science in this new dispensation? Will 'research', as some cynics suggest, consist largely of commissioned studies effected by house-trained 'consultants'? Much of the LIC's research was stated to be policy driven but *some* funding was to provide for what was explicitly 'blue skies' research, whatever that might mean.

Some academic researchers look to the Joint Information Systems Committee (JISC) of the Higher Education Councils for salvation. Certainly JISC has been significant in recent years in promoting and funding information development work, through its Electronic Libraries Programme (eLib) and other initiatives. But JISC is not a research council and does not support research as such, only those investigations relevant to its overall development programme. Various research councils have been an adventitious source of information research funding in the past, but only when they could be persuaded, on relatively rare occasions, that the proposed research might benefit their overall disciplinary mission. Any reference to libraries or information services in an application was likely to invoke John Martyn's axiom: 'A proposer should assume that, given the slightest opportunity to refer his application to another body, the funding agency he has approached will take that opportunity.' Fortunately there is a relatively new research body – the Arts and Humanities Research Board – that will entertain proposals about LIS research; unfortunately the funds that it has available (for LIS) do not match those of the heady days of BLRDD.[1]

Garfield's proposals do not infer the ditching of science and its replacement by technology. Indeed, 'the theory of information science is an essential underpinning to our professional work but information technology provides feedback and nurtures theory'. This is a nicely balanced, realistic approach. Here in the UK we have the MLAC 'which will replace both the LIC and the Museums & Galleries Commission no later than April 2000' and will presumably continue the LIC's function which 'substantially [sic] expanded its research role by assuming responsibility for the research commissioning and funding activities formerly carried out by the British Library' [3]. As a member of an ageing and inconsequential Jacobite tendency, I concede that things have inevitably changed since the salad days of DSIR. But I am not persuaded that there has been much progress.

References

1. Inside ASIS. President-elect proposes name change. *Bulletin ASIS*, June/July 1999, 2.

2. Yes vote for unification with LA. *Inform*, 218, October 1999, 1.

3. Press release. Http//www.lic.gov.uk/publications/pressreleases/me0908.html

Notes

1. Getting one's hands into what American politicians would call the 'pork barrel' of the European Commission has, of course, proved lucrative for several academic researchers. But the machinations required for that kind of support are outside the scope of this editorial.

Tales on the Internet: making it up as you go along

Chris Frost
School of Journalism, University of Central Lancashire, Preston PR1 2HE
c.p.frost@uclan.ac.uk

The Internet's email system is a fast and efficient communication method and one that has become particularly popular for rumours and hoaxes. These include virus alerts, one of the favourite types of hoax, and urban myths or contemporary legends.

An interesting element of these types of rumours and hoaxes is:
- the apparent reduction in caution about their retransmission;
- the speed and ease of re-transmission that email offers;
- the extensive detail that is often included at variance to standard word of mouth transmission of rumour.

This paper examines a single case study in order to investigate the differences between e-rumours and word of mouth rumours. Theory suggests that detail adds verisimilitude to a rumour and this could certainly explain why such emails often provide extensive detail. The email system eases the job of re-transmitting and this might explain why readers are prepared to broadcast such rumours without further checking.

Introduction

The Net is thought by many to be the method of mass communication of the future. Predictions that it will overtake newspapers and even traditional broadcast methods can be found in many journals. There is also evidence that personal communication is set to be usurped by the Net. Emails are now the standard communication method for many academics and the business world is quickly catching up. It is clear that within five years, email will be a major method of communication even at a personal level. Since the start of the Web, there has been a growth in the number of rumours and hoaxes transmitted by email.

For the purpose of this paper a hoax is a rumour started deliberately for malicious reasons or as a practical joke. A rumour is unsubstantiated hearsay. These hoaxes/rumours, which include virus alerts – an increasingly common form of hoax that is unique to the Internet, are significantly different in character to the types of rumour carried by word of mouth.

While the Net widens the community of interest and allows rumours and gossip to flow across international borders it is only recently that the people using the Net have become sufficiently diverse to cease being one specific community. Although, even now, there are common areas of interest to all Net users.

There are added advantages of the transmission of rumour by email. The Chinese whisper effect of serial transmission, which several psychologists cite as the main reason for the addition of distortion, combined with levelling and sharpening of the rumour rarely happens with emails. It is easy to pass the message on in exactly the same form as it arrives. Nonetheless, different versions of rumours and hoaxes do circulate and unless we assume that someone was told about the story and that person then re-transmitted onto the Net, the only explanation for changes in a rumour is that someone wanted to insert those changes – either because they have extra information or because they wish to add to or change the hoax.

This paper intends examining the difference using email, as the transmission system, brings to the type of rumour and its transmission and compares these differences with spoken rumours. A case study of a modern e-rumour will be used to illustrate the differences.

Literature review

Rumours have been an area of serious study for psychologists since the turn of the century. Despite this, there is not much published about them. Allport and Postman wrote *Psychology of Rumor* in 1947 in the shadow of the war in which rumour had been a major problem. They identified the two main elements of a rumour as being its *importance* and its *ambiguity* [1, p. 33). By this they meant the importance of the rumour to the listener and the extent of the ambiguity of the event being described. They explained that a rumour about events which were not of importance to the listener were not passed on while if a listener had confirmed details about an event, those were not passed on either. So if a rumour were told to a New Yorker about how a small Yorkshire brewery put body in its beer by adding dead rats, he probably would not pass it on (unless he was mischievous and altered the rumour to a small New York brewery). Whilst a Yorkshireman might pass on the tale, the head of the brewery would not – he knows dead rats are not used and so the tale is not ambiguous, just wrong.

Allport and Postman used these two elements to produce a formula for the intensity of rumour. They posited that R~IxA; that rumours grew in inverse proportion to the importance factored by the ambiguity [1, p. 33]. If the story was not ambiguous, then the factoring totalled zero and there was no rumour, the same applies if the rumour is of zero importance. Only if the subject of the rumour is of some importance and some ambiguity to the listener, does the rumour spread. Allport and Postman also talked about the rumour process 'levelling' and 'sharpening' rumours [1, p. 75]. By this they meant that some detail was always levelled out in an effort to make the story more concise whilst the most important details were sharpened or made more prominent. They also talked about 'assimilation', the tendency of reports to become more coherent and consistent with the beliefs of the listeners [1, p. 101]. However, these effects were defined in their study of serial transmission. Their experiments involved showing a picture to an observer who then had to describe the picture to another person, who then had to describe the picture to the next subject and so on. Eventually after about seven or eight transmissions the story become considerably changed, with marked levelling and sharpening.

Allport and Postman did most of their studies on serial transmission, which was, they believed, how rumours worked. Tamotsu Shibutani, however, believes that rumour is not a serial process but is a collective group activity [2, p. 9]. Using a series of case studies he shows that there are several players in the development of a rumour and its successful transmission and that rumours are a reflection of people's fears and insecurities or very occasionally (and he believes it is extremely rare) malice. Shibutani says that unverified reports are considered seriously only in the absence of reliable news. Allport and Postman also found that a *trusted* source of information could kill a rumour dead [1], either by confirming or refuting it. Allport and Postman discuss at length the role rumour clinics played in US newspapers during the war. These were columns in many US newspapers which discussed a rumour doing the rounds, analysing it and effectively refuting it by providing sound evidence from a trusted source.

Shibutani adds some other rumour effects. He believes that rumours are subject to a 'plausibility' test [2, p. 77]. Implausible rumours are either adjusted to make them more plausible or ignored altogether. 'The definition most likely to be accepted is one that fits in with presuppositions held more or less in common within the public [2, p. 77].' Shibutani sees rumour development in normal times as a process taking place where the 'unsatisfied demand for news is moderate, collective excitement is mild and rumour construction takes place through critical deliberation [2, p. 94].'

The argument for importance being a

major factor in rumour transmission is supported by Anthony [3, p. 91]. In an experiment she found that two groups of 'chronically high anxious' high school students transmitted a rumour almost completely whilst two groups of 'low anxious' high school students had very little transmission of the same rumour. She argues that this suggests that anxiety is one of the factors in rumour transmission. She also discovered that in all test groups the rumour was passed to at least one other person [3, p. 97] showing that it is the anxiety of the listener, not that of the teller, that is significant. She also supports the view that importance is another factor in rumour transmission [3, p. 97].

Rumours are presented in a number of different ways. Allport and Postman see them as serially transmitted with all the opportunities for distortion that this brings. Shibutani says they are group activities with a number of cooks adding their ingredient to the brew. Rosnow and Fine believe that form and function of a rumour can be highly varied '... rumor may refer to a group effort to define what is occurring, or a rumour may be a packaged story, deliberately planted [4, p. 22].'

Method

Case studies are a good way to examine the use of email and the Web for rumour and hoax transmission. Many examples are available and their transmission over a short period can be examined. This observation can be supported by interviewing key players in the transmission process, although it is worth noting that those who will allow themselves to be interviewed seem to be well-intentioned innocents who took the message at face value. Other research methods would not lead to anything like as much useable data. The case study chosen was the hoax reprinted below concerning the use of mobile phones and is transmitted by an email. This allows for easy, quick transmission to large numbers of people using address books, where a message can be sent electronically to all the people on the list simultaneously.

This email has been circulating the Net for a number of years (AT&T website July 1999) [5]. It was sent to a police officer in Guildford, UK, who in good faith sent it on to a number of people as part of a crime prevention initiative including, finally, a member of staff at the University of Central Lancashire, who circulated it to colleagues in the University at which point it was picked up for use in this study. The names of the people concerned have not been used (although they are known and their unwitting involvement has been confirmed) to spare them embarrassment. By the time the rumour's circulation in the University was noted and the first colleague had been emailed to warn her this was a hoax, she had received a further six emails from other colleagues within the institution detailing the same rumour.

The email

> 'Hot off the press from WPC S________ at Guildford Police Station: If you get a phone call on your mobile from a Cellnet or Vodafone engineer, telling you that they are doing a check on your phone and that you have to press #90 or 09#. END THE CALL IMMEDIATELY. There is a fraud company on the go who have devised a device that once you have pressed #90/09#, they can access your SIM CARD and make calls from it at your expense. Tell as many people as you know that have mobile phones to put this to an end.
> Regards'

The first thing to make clear is that the rumour is almost certainly not true. The press office at Vodafone has confirmed that the company was obliged to put out statements denying the rumour:

> 'There has been some concern about a potential fraud that can be incurred by users of mobile phones by dialling 90#. Vodafone has been aware of these claims for some time and has made lengthy investigations but has been unable to recreate the effect or otherwise substantiate the claims. In consultation with other networks, we believe this to be a false alert and an unsubstantiated claim.'
> (Vodafone group services press office, July 1999)

The Surrey Police at Guildford were also obliged to deny the story:

> 'Surrey police received a report that mobile phone users were being targeted by conmen pretending to be engineers, who asked them to key in a sequence of digits and then managed to bill subsequent calls of their own to the victim's account. The force passed on this information to the public in good faith as part of a crime prevention message. After consultation with BT Cellnet and Vodafone, who have carried out extensive tests, it is now believed that the suggested means of deception will not affect the equipment in the way described. Mobile phone users are not believed to be at risk from this kind of fraud.'
> (Surrey Police press statement, July 1999)

The background to the story is more detailed. It seems there might be a germ of truth in the hoax. The main story has been circulating the Internet for some time. AT&T the US telecoms giant warns on its website that this fraud can only apply to businesses with switchboards and then only in certain very specific circumstances. AT&T's website says:

> 'This type of fraud involves a perpetrator who calls an office and cons an unsuspecting worker into transferring him or her to an outside line. The perpetrator then starts dialling calls that are charged to the owner of the PBX. In this latest version, the caller claims to be an AT&T service technician 'repairing' the phone lines and convinces the recipient of the call to help out by transferring him to an outside line and then hanging up [5].'

The hoax has probably been circulating the Net for some years, but according to the urban legends page of the miningco website, an incident happened on Saturday, 24 January 1998, at a Naval Air Reserve Base in New Orleans:

> 'Quarterdeck received a telephone call from an individual identifying himself as an AT&T Service Technician that was running a test on our telephone lines. He stated that to complete the test the QMOW should touch nine (9), zero (0), pound sign (#) and hang up. Luckily, the QMOW was suspicious and refused.'

So it seems that in the US there may be a small amount of truth in the rumour for land lines:

> '. . . just to be anal-retentive, let me say one more time that, unless you have to dial 9 to get an outside line at home, this scam does not affect residential telephone users. [It also probably does not affect non-US telephone users. This is especially true for British telephone users whose telephone system is so complex that NO ONE in the UK knows how to use BT's phones (although I am sure that BT users are currently dealing with some sort of 'dial q-seven-pi-cromwell-eleventeen-tomato' scam)] [6].'

Despite all this, warnings about the scam were still posted on the Royal Canadian Mounted Police site in July 1999 [7] as genuine.

Discussion

Laws of Rumour

The limited literature available on rumour seems to suggest that there are three main elements to consider:

- importance of the rumour to the listener;
- ambiguity of the rumour;
- anxiety of the listener.

How does the mobile phone rumour fit these elements?

Importance

The mobile phone rumour is likely to be of moderate to considerable importance to mobile phone users who would be concerned about a fraudulent use of their phone. Whilst no research has been found showing a link between Internet use and mobile phone ownership, Internet users are probably more likely to possess a mobile phone or aspire to possessing a mobile than most other groups in society. They are technologically literate and are either in work (and using the

Internet there) or have sufficient funds to buy a computer and use the Internet at home. Even students, who are often heavy Internet users at their educational institute, own mobile phones or aspire to use them for the same reasons.

Plausibility

Shibutani says that rumours become more plausible during transmission. He gives examples of rumour that seem bizarre to outsiders, but because of the internal social forces may be entirely plausible [2, p. 80]. In this example, though, the story is entirely plausible. Most people are aware that phone fraud was a considerable problem with the old analogue mobile phone system. It therefore sounds entirely probable that a scam allowing criminals fraudulent access to a phone's SIM card could be discovered. Most people are not technically able to decide whether this is possible. Interestingly, even those who are technically able cannot guarantee that a way cannot be found to do this. They talk about it being unlikely or unusual. The Vodafone warning says 'has been unable to recreate the effect or otherwise substantiate the claims', rather than deny them completely. It is this difficulty the technicians have in denying absolutely that something could happen that helps make the rumour believable to the non-technically minded.

Ambiguity

Ambiguity is an important element of any rumour. The ambiguity of the rumour as reproduced here is minimal. This might normally lead to the rumour not being passed on according to Allport and Postman. They give the example of a man who has won $50,000 on the lottery. He is not likely to pass on a rumour that he has won $1m because he knows the truth. This hoax is no 'attempt to provide structure in an uncertain situation' [4, p. 28] it seems to be a deliberate attempt to deceive.

In the specific example, the very lack of ambiguity and good use of detail adds verisimilitude. It sounds plausible, the technical detail panders to the general public's knowledge that codes allow us access to SIM cards and that therefore there might be secret codes to allow others access.

Trusted source

To cap it all, in this instance (although remember, this was just a happy coincidence that strengthened this version of the rumour) it is supported by an excellent source – a police officer. Who else should know more about the latest crime methods? It is this element of adding a trusted source that is often significant in such hoaxes. Warnings of this sort are always supported by authoritative sounding sources. In this instance the police failed to check before transmission and have added to the problem. Their role here was not welcomed by Vodafone (see press statement).

Internet effect on rumour transmission

An area of difference between electronic transmission and word of mouth transmission is that whilst in word of mouth the rumour will be sharpened so that only the relevant details are transmitted, in electronic transmission this does not happen. The email is already sharply crafted to limit it to the minimum detail required with the maximum apparent validity.

Allport and Postman say that the ambiguous message, which cannot be verified, is likely to be passed on in the nature of a rumour. But with this story there is no revalidation required. Nor is the listener likely to forget or distort the detail. Forwarding an email keeps the exact wording intact and allows it to be sent to large numbers of people on email lists. Email technology allows the rumour to be transmitted to scores, possibly hundreds of people, simultaneously with the use of address group lists.

There is also something about the Internet itself that actually seems to enhance reliability. It may be that because the hoax pages are usually produced to the same high standards as any other page, it is difficult to tell the difference. We are used to equating high production values with high credibility, although there is no good reason why this should be so. This connection between high quality display or presentation and trustworthiness has long been used by conmen and tricksters.

David Emery comments on the reliability of Internet material in an Internet essay:

> 'When I forward an email message to you, I cause an exact duplicate of a piece of information I've received to appear in your mailbox. I do not communicate the information to you, I simply transfer it. The information is thus autonomous – 'untouched by human hands', as it were – which, by the curious logic of hyperreality, actually seems to heighten its air of authenticity. It also negates the concept of accountability, because even though I can pass the same piece of information on to you and a hundred other people by clicking 'Forward', I am in no practical sense its 'source' [8].'

The harm test

It seems probable that faced with this rumour, receivers consider it of interest and importance. Do they have any reason to doubt it? The original source sounds genuine, as does the message, while the message forwarder (who as Emery points out is not considered to be the source) is usually known to and possibly trusted by, the receiver. This may be why authenticity seems to be added – we know the transmitter and apply the credibility we give them to the message, which they have transmitted without any attempt at authentication.

It is also a no harm rumour. If no one ever rings you on your mobile and asks you to type in the code numbers, then there is no harm in passing on the rumour. (Although the press office at Vodafone might not see it that way. This email caused them a lot of work.) Only if the engineers really do need a mobile phone user to key in that number for proper engineering reasons and is refused, is there likely to be harm. Since the passer of the rumour is unable to test the rumour, but can see no harm in transmitting it, there can be no social or moral imperative to prevent transmission.

Conclusion

This story fulfils all the criteria of a good hoax/rumour. It is important to its readers. It seems supported by good sources and is not ambiguous. Its transmission methods preclude sharpening and levelling and ensures there is no ambiguity. It is important enough to pass on. There is enough detail to give it the ring of truth and the subject matter and the immediate source (the transmitter), and possibly the original source, seem to be trustworthy and there is no apparent social or moral impediment to transmission. This type of email shares common ground with a number of virus hoaxes and, like them, are likely to be circulating the Internet for some time.

References

1. Allport, G. and Postman, L. *The psychology of rumor.* Revised ed. New York: Holt, Rinehart and Winston, 1965.

2. Shibutani, T. *Improvised news.* Indianapolis: Bobbs-Merrill, 1966.

3. Anthony, S. Anxiety and rumor. *Journal of Social Psychology,* 40(3), 1973, 597–620.

4. Rosnow, G. and Fine, R.L. *Rumor and gossip: the social psychology of hearsay.* New York: Elsevier, 1976.

5. AT&T. http://www.att.com/features/0398/90pound.html (visited July 1999).

6. Crispin, P. 1999. http://urbanlegends.about.com/library/weekly/aa021898.htm

7. Royal Canadian Mounted Police. 1999. http://www.rcmp-grc.gc.ca/html/scams.htm

8. Emery, D. 1999. http://urbanlegends.about.com/library/weekly/aa081097.htm

A strategic approach to research using Internet tools and resources

Lyn Robinson
School of Library, Archive and Information Studies, University College, Gower Street, London WC1E 6BT
Lynxi2000@cs.com

This paper describes a method for carrying out research in a multimedia environment encompassing printed, electronic and Internet based resources. The framework was used to compile a listing of toxicology information resources, as part of a larger study on the effect of new information and communication technologies on scientific communication within this field. It is likely, however, that the approach can be applied to research in any subject.

Introduction

A research query is open-ended, and must be continued until, in the judgement of the researcher, a satisfactory conclusion has been reached. This is in contrast to a 'reference query', in which a single answer is sought, and the process may be ended when it is found. Research queries require the creation of a list of resources on which to base the investigation. It may be that such a listing is the end point of the research, which may be, for example, the compilation of a subject guide or directory. Items on the list may be in printed or electronic format, and increasingly the latter type may be available via the Internet. As the Internet does not, and may never, offer access to every resource, it is essential to consider an approach to research which includes other resource types within its structure. In addition, resources in various formats may each be appropriate for varying circumstances, as described by Gilster [1], with regard to the choice between electronic and printed books.

Unless the resources are already known to the compiler, a research query must involve some searching to identify information that will answer it. Because of the proliferation of sources, particularly where the Internet is concerned, the problems will almost certainly lie in the selection and organisation of a large amount of material, rather than a difficulty in finding anything at all. This will require an evaluation step, in which resources, which have been identified as relevant to the subject, are assessed against a pre-defined set of criteria, to decide whether they are appropriate for inclusion. A further, and initial, stage is usually required; the construction of an intellectual framework within which to fit the material found, and thereby to make sense of it. Assuming that the compiler has some knowledge of the subject area, and of the kinds of resources likely to be found, then this framework will be established at the outset; otherwise it must be done as the process progresses, by considering material found from initial searches.

Creating such a framework is equivalent to setting a context for the resources found, and is a part of the 'critical thinking' approach, often proposed as an essential skill in making effective use of information resources in general, and the Internet in particular. Gibson and Meade [2, p. 144] express this idea clearly:

> 'Context mysteries for Internet information and searching will continue to be a problem. The critical thinker will have to develop a site orientation to Internet resources in order to invest them with meaning. Retrievals in cyberspace have to be placed in some kind of mental framework created by the searcher. Various kinds of mental frameworks can be created, including classifying Internet resources by origin

> or provenance, type of data, need for special manipulation or file management, or even an intellectual seriousness sale that would presumably locate certain Usenet groups at the low end of the scale while placing the Oxford Text Archives or the Coombspapers archive at the high end of that same scale. Critical thinkers will invent frameworks that help create context in what often seems like a context-free environment.'

This reference to a context-free environment reminds us that the Internet poses particular problems in this respect, precisely because of the consistent presentation, by a web browser, of information from a variety of origins. The visual and tactile, as well as intellectual, context clues present in the examination, for example, of books and journals in a paper-based library are entirely lacking in the digital environment.

Similar arguments are made by the OMNI (Organising Medical Networked Information) group:

> 'Establishing the context of a resource can be a challenging task. By context we refer to aspects such as scope, audience, authority and provenance. Networked information resources tend not to have a common set of features such as statement of responsibility, introduction or preface from which to draw such information. The policy of pointing directly to individual resources places a greater responsibility on [list compilers] to examine sources of information about a resource (e.g. home pages, parent documents and help files), in order to place a candidate resource in context [3].'

A methodology for creating a resource listing, therefore consists of the following three stages:

- choose/create a framework;
- identify resources;
- evaluate resources.

Each of these will now be examined in more detail, and a specific methodology for the creation of a toxicology resource listing will be outlined as an example.

Creating a framework

Existing subject guides and resource listings, for toxicology as well as for a variety of other subject areas, make use of a variety of schemes and frameworks for categorising resources, and thereby giving structure to the list.

Since the list can be created and maintained in digital form, it might be argued that choice of a framework is not so crucial, since the electronic format means that multiple frameworks could be maintained, and resources selected *post hoc*, by whatever criteria seemed appropriate. However, this would be to miss the main point of the establishment and use of such a framework. It acts as an intellectual aid to the creator of the listing, in helping to impose structure and context on the mass of material likely to be encountered during the compilation of the list, thereby helping to ensure consistency, balance and common sense in the choice of material. As will be seen later, if a framework based on type of material is chosen, it acts a powerful aid to the compiler, in enabling a 'cascade' style of resource identification. Subsequently, a suitable framework should also act as an aid to the user of the list, acting as a guide to the nature, and interrelations, of what is provided. This is not to say that supplementary access points to the main framework cannot be provided for convenience, but this does not detract from the necessity of the main context-setting framework.

There are four forms of general framework categorisation, excepting those, exemplified above by Gibson and Meade, which are Internet specific, that appear feasible for a subject such as toxicology:

- by subject;
- by format;
- by location;
- by type of material.

The first of these, a subject categorisation – in our example dividing material by specific type of toxicology – is commonly used for subject guides, including some in the toxicology area. Indication of subject is obviously essential, but this form of categorisation is not suitable as a main framework for this list.

For one thing, many resources cover multiple subjects, leading to duplication in listing and possible confusion. More importantly, a primary breakdown by subject would bring together very diverse forms of information, and would be less than helpful to the user of a list of this sort, while it might well be appropriate for a list involving only a single format of information, e.g. online databases, or printed journals.

The second form of framework is one based on format of information: printed book, online bibliographic database, web page etc. While this is a useful element of information to have in the description of each resource, it is not suitable as the primary framework. For one thing, some resources exist in several forms – e.g. Index Medicus/Medline exists as print, online, CD-Rom, ... – leading to duplication. For another, it would be rare for a user to look for specific formats, unqualified by other considerations.

The third form of categorisation divides resources in the list according to country of origin. This is done, to some extent, in Wexler's 1988 [4] listing of toxicology resources, which is the most comprehensive listing to date. There is some justification for this in the case of legislative and regulatory aspects, which may well differ from country to country, but it is not overall a particularly helpful approach. This is particularly so when a good deal of information is networked, and it is largely irrelevant to know, and may be impossible to determine, the geographical location of a particular resource.

This leaves the fourth type of framework, in which resources are initially categorised according to their type, in general terms. This seems particularly suitable for a list of this sort – assuming that it will be supported by other descriptive elements including subject indexing – since it allows compiler and user to approach the issue – the research topic – with an initial question as to what 'sort' of information, and hence, resources may be available. It leaves open the question of exactly how the categories will be defined.

Consideration of possible frameworks of this sort (based on examination of a variety of examples of subject guides, backed up by searching of relevant bibliographic databases and of the Internet) leads to the conclusion that the most generally useful one is an adaptation of a 'traditional' library/information framework, based on structure of literature, particularly scientific literature. Resources are divided into primary, secondary, tertiary, and (sometimes) quaternary categories, essentially according to their distance along the communication chain from the original 'raw' information. A particularly clear and full description is given by Grogan [5]. Adaptation is necessary because this model was derived for a system of publication based purely on printed products. Limited modification coped with the inclusion of computerised sources, e.g. machine-readable abstracting and indexing services, but the arrival of the new forms of information product on the Internet require the model to be extended further. Nonetheless it is noteworthy that this relatively elderly model still serves very well in helping to clarify the allegedly chaotic system of (largely networked) information of the present time.

Under the adaptation of this system, resources are categorised as follows:

Primary: the original information, in whatever form it takes:

- journal articles, pre-prints (paper and electronic);
- reports;
- data tabulations;
- laboratory notebooks;
- diaries;
- memoranda, letters, email messages, postings to newsgroups;
- conference proceedings;
- theses, dissertations;
- patents;
- standards;
- regulations and legislation;
- trade and product information, technical manuals;
- course notes, interactive training materials, syllabi;
- company reports, financial data;
- organisations' and individuals' homepages.

Secondary: 'worked over' knowledge, which

organises the primary material in some way:
- reviews and summaries of progress;
- indexing and abstracting services;
- data compilations, databanks;
- monographs, textbooks, treatises;
- reference works, encyclopaedias, handbooks, data tables;
- dictionaries, thesauri, classification schemes, glossaries;
- pharmacopoeias, formularies.

Tertiary: rather diverse sources, which aid the use of resources at other levels; typically they do not carry 'subject' information at all:
- yearbooks, directories;
- bibliographies, reading lists, location lists of periodicals, lists of abstracting and indexing services, library catalogues;
- literature guides, lists of 'recommended' or 'core' resources;
- virtual libraries, subject gateways, specialist search engines, annotated lists of internet sites;
- subject listings of mailing lists and newsgroups;
- listing of research in progress, indexes of expertise;
- guides to libraries, online host systems, other sources of information;
- organisations;
- listings of academic departments and courses.

Quaternary: give access to resource listings at a high level, not usually subject specific:
- lists of lists;
- bibliographies of bibliographies;
- listings of library catalogues and organisations.

Primary and secondary resources are subject specific by this definition; tertiary may or may not be. Quaternary are not usually subject specific.

There is a further, zeroth order, of information; the 'raw' data which comes, for example, from laboratory instruments: however, although this can now be readily captured, stored and transmitted elsewhere, it is not truly an information resource until it has been organised and described so as to be understandable in 'public use', at which stage it can be regarded as primary information.

Identifying the resources

Identifying the resources to be put into a list of this sort will inevitably be, to some extent, a subjective process, depending on the subject knowledge and experience of the compiler. However, if the framework suggested above is used, this will provide an outline procedure to be followed, as an aid to making the procedure systematic.

The basic principle is to begin with the higher level resources, and use them sequentially – following listings, references, hypertext links etc. – to find those at lower levels. Thus, quaternary resources are used to find tertiary, tertiary to find secondary, secondary to find primary. In principle, in a solely print-based communication system, this approach could be followed in its 'pure' form, without any return from lower levels to higher; though even here there could well be exceptions, a reference from a journal article (primary) to a review article (secondary), for example. When one is dealing with resources in hypertext form, links 'upwards' will frequently be encountered, so that the compiler must expect some iterations to higher levels. Even so, the general principle of starting with resources at higher levels and using them to find those at lower levels is still a useful means of attaining a systematic procedure.

There is then the question of *how* resources are identified at each level. There are, again in general terms, three ways:
- personal knowledge;
- expert advice;
- searching.

Personal knowledge: under normal circumstances, the compiler of a resource list will have some subject expertise, and therefore some knowledge of resources. Indeed, there is a recent trend for such expert knowledge to be the sole means of input to resource listings [6]. While this seems less than ideal, if anything approaching a comprehensive and unbiased list is required (there is no such thing as a personal encyclopaedic knowledge of sources in the Internet age), it would be perverse to fail to make use of the expertise of the compiler. This will therefore normally be the first source of entries to the list at every level. (The only exception would be when a research query of this form is handed to an information specialist, who may have no personal knowledge of the area. This first stage would, perforce, have to be omitted, and the procedure would be the poorer thereby.)

Expert advice: this implies advice of subject experts above and beyond that of the compiler. At least one toxicology resource listing was produced using this method alone, by asking a panel of experts for their favoured resources, and consolidating the results [7]. However, though arguably preferable to accepting the views of a single compiler, it is not a method that can be used alone. It is likely to be used in conjunction with other methods, as either the initial or the final step in the procedure. It would be appropriate for it to be the initial step in the case where the compiler had no personal expert knowledge, and this was used in its stead; the 'reference interview' carried out by an information specialist before carrying out a search on behalf of a client is an analogy.

More usually, and certainly in the toxicology study, this step comes late in the procedure; expert advice is sought as a check on, and a means of seeking additions to, an initial list compiled on the basis of steps one (personal knowledge) and three (searching).

The seeking of expert advice has usually been thought of as being carried out on a personal, perhaps face-to-face, basis, and this may well be the case. However, electronic communications could certainly be used to expedite and widen the process; perhaps by submitting a list for comment from members of a relevant electronic mailing list or discussion forum.

Searching: the third means of identifying items for the list is searching. This rather general term covers a variety of processes, which will vary according to the subject area of the list, and also to the type of material which will be included in it. There will generally be specific searching options for different levels of material, and also for material in different formats. However, there will be some overlap in the latter: Internet search engines, for example, will, by definition, return web pages, but some of these may contain information on printed books and journals, or on means of accessing online databases.

Searches are therefore carried out using tools appropriate for the subject matter and the type, or level, of material being sought. In general, this will mean using Internet search tools and bibliographic databases to identify material at each of the four levels. This requires that the compiler be able to identify suitable search tools. To some extent this is a circular argument, since these tools may well be secondary or tertiary resources to be included on the list. An example here would be subject specific bibliographic databases such as Toxline, which will be both secondary resources for toxicology information, and hence to be included in the resource list, and also sources to be searched for items describing toxicology information provision. If the compiler has some initial knowledge of sources, they will probably know of such databases. If not, the initial searches will have to be restricted to more general sources (e.g. LISA), which will hopefully identify subject specific databases, to be searched as a second phase.

Three kinds of bibliographic database will be useful for identifying resources for the toxicology resource list:

- Subject specific databases, which may include items on information resources in their area. For toxicology, examples would be Toxline, Medline, and Biosis.
- Databases covering library/information literature, which would be likely to include items describing subject resources: e.g. Library and Information Science Abstracts (LISA), Social Scisearch.
- 'General' databases, covering material in a particular form: e.g. Books in Print, Dissertation Abstracts, and NTIS (for governmental reports, particularly from the USA).

Internet search tools include search engines and directories. Since the number of these is high, and their nature changes so rapidly, it is inappropriate to be too specific as to which to use at any one time, but selection should be based on coverage of Internet material and frequency of update, as suggested by Sullivan [8], as well as any which may cover the relevant subject particularly well, e.g. the environmental organisation directory at www.webdirectory.com. (This is again a circular argument, since subject specific search engines would be useful resources in themselves to be identified and added to the list.) Directories emphasising 'professional' material, such as BUBL (www.bubl.ac.uk) would also be favoured.

It will also be necessary to choose search terms in order to find relevant material from these sources. This is a matter for the expertise of the compiler, but in general it will be necessary to use terms reflecting the subject, i.e. in this case 'toxicology', with those reflecting the concept of 'information resources', and any more specific concepts suggested by the framework, e.g. 'journals', 'library services', 'databanks', 'training materials', 'mailing lists', 'newsgroups'. In general, very simple searching, perhaps just the subject indication, can be used in general and library/information sources, especially when quaternary resources are required. For the subject specific resources, and for primary sources, more detailed search specification will usually be needed, because of the larger amount of potentially relevant material that is likely to be retrieved.

This straightforward approach to search tool and term selection will usually suffice for the identification of material at quaternary and tertiary levels. To identify secondary resources, more level specific sources will usually be added to the bibliographic databases and Internet search tools chosen. This could include the documentation of providers of online and CD-Rom services, which will enable identification of relevant databases and databanks. At primary level, additional specific sources will also be searched, e.g. library catalogues. It may also be appropriate at this stage to examine physical collections of material, or items describing them, e.g. collections of books and journals in specialist libraries. Further details describing the construction of a list of toxicology resources are given in Table 1.

Evaluating and selecting resources

The process described above can generate large numbers of potentially relevant material. As will be clear, the identification process itself, to some extent, will provide particularly 'useful' or 'significant' material, if such exists. Nonetheless, there remains the need for the list compiler to make choices from among the material identified, or at least to ratify its appropriateness for inclusion. While there will inevitably be some element of subjective judgement in this selection/ratification, and different list compilers might be expected to make some different decisions, an element of consistency can be imposed by the use of selection guidelines, if not rigid criteria.

Guidelines for selection of material have always existed in library/information work, originally under the heading of 'book selection'. The topic has been given impetus recently in the context of evaluation of Internet information, in particular by the establishment of 'quality gateways' such as OMNI and SOSIG (Social Science Information Gateway) (see, for a biomedical perspective, Anagnostelis [9], and more generally Tate and Alexander [10]. While the precise sets of criteria used for these gateways, which are restricted to listing Internet sites, are too restrictive for our toxicology list, which aims to cover relevant items in all formats, they are a good basis for derivation of a more general set of criteria.

The most relevant set of criteria for initial consideration for a list of toxicology resources is the set of criteria used by OMNI, since these have been designed specifically to deal with biomedical material.

OMNI's criteria are set out below in brief (a full version can be found at OMNI [3]):

- Scope (subject area, breadth, depth, time period and format or type of information).
- Intended audience.
- Authority (who are the providers?).
- Provenance (how long has a resource been

Table 1. Detailed procedure for compilation of a list of toxicology resources

construct framework, based on type of material, specifying any particularly important categories

identify a set of bibliographic databases and Internet search tools to be used (see Notes)

decide on search terms and strategies to be used (see Notes)

identify level specific resources to be used (see Notes)

identify quaternary sources
from personal knowledge
- search Internet search tools
- search bibliographic databases

identify tertiary sources
from personal knowledge
from quaternary sources already found
- search Internet search tools
- search bibliographic databases
- search level specific sources as needed to identify:
  library and information services
  toxicology societies
  online host systems etc.

if new quaternary sources are found, check for new tertiary sources in them

identify secondary sources
from personal knowledge
from references/links in tertiary sources already found
- search Internet search tools
- search bibliographic databases
- identify relevant online hosts/CD-Rom providers – search their documentation and use cross-index search to identify useful databases
- search any other level specific entities

if new tertiary sources are found, check for new secondary sources in them

identify primary sources
from personal knowledge
from references/links in tertiary sources already found
from references/links in secondary sources already found
- search Internet search tools
- search for mailing lists and newsgroups
- search bibliographic databases
- search library catalogues and 'books in print files' for books
- search periodicals databases for journals
- look at stock of specialist libraries for books and journals

if new tertiary sources are found, check for new primary or secondary sources in them
if new secondary sources are found, check for new primary sources in them

Recheck all entries against evaluation criteria.

Submit to independent subject experts for checking.

Check final list for overall balance.

available, either in print or as an electronic version?) How is it maintained?
- Coverage.
- Accuracy of information content.
- Currency/frequency and regularity of updating.
- Uniqueness/comparison with other sources.
- Accessibility and usability.
- Charging policy.
- Special requirements (hardware and software).

- Software reliability.
- Copyright.
- Language.
- Design and layout/user interface.
- User support/documentation.

A generalised and modified version of this list can provide a set of criteria for the selection of material for a list of toxicology resources:

- Authority: who produces the resource; what qualifications have they to do so?
- Provenance: in essence the stability and longevity of the resource.
- Objectivity, evaluation: is the information evaluated and balanced; or does it present a single point of view?
- Currency: how up-to-date is the resource; how often is it up-dated?
- Accuracy: how valid is the information; what is the 'quality control'; how are errors corrected?
- Coverage, completeness: how complete, within its stated area of coverage?
- Relevance: how useful would the resource be, and to what sort of users?

These are, in fact, quite similar to the five 'classic' criteria – accuracy, authority, objectivity, currency, coverage – which Tate and Alexander [10] remind us were traditionally used to evaluate print resources. Provenance has been added to reflect the more dynamic networked environment, and relevance because of the subject specialist nature of this resource evaluation.

In practice, these criteria can be applied in a context sensitive way. If the resource is being included as a 'significant' resource, then it can be judged more strictly than if it is included as an example. In the latter case, the criteria would simply be 'how representative is this of material of this kind?' In those cases where the resource is the only exemplar of its kind, then that is sufficient criterion for inclusion.

Since the toxicology listing is intended to provide an illustration and exemplification of the kinds of toxicological resources available, rather than to give a set of 'quality assured' resources, some of the restrictions applied by OMNI are not relevant here. For example:

> 'Resources are included only if they contain substantive information ... Personal pages or simply collections of pointers to other resources do not meet this requirement and are de facto excluded... Ephemeral material, such as advertisements and e-mail messages or newsgroup postings, is also normally excluded [3].'

Procedure for construction of resource list

In summary, the outline procedure for identifying resources for list construction is as follows, assuming that:

- The scope of the subject is understood.
- The list compiler has some subject expertise.
- The compiler has an initial knowledge of some sources.
- A framework for setting context is constructed.
- The investigator's own expert knowledge is used to create a proto-list of resources at all four levels.
- Resources on this proto-list are examined, to suggest further resources.
- Searching tools (bibliographic databases and Internet), and search terms, are identified.
- Resources are searched for at each of the four levels in turn, starting at the highest, and using searching resources appropriate to each level. Resources found at each level are used to identify those at lower levels. Where a resource gives a reference (in printed or bibliographic files) or a link (in HTML files) up to a previously unidentified resource at a higher level, this is checked to see if it led to a further lower level resource.
- The resources found are evaluated against the appropriate criteria and entered into the list if appropriate.
- The resulting list is submitted for checking by independent subject experts; any new resources suggested by them which are judged acceptable are entered into the list, and used as starting points for a check to identify further resources.
- The final list is checked by the investigator

for balance, consistency and common sense.

Conclusion

The explosive increase in information resources realised by the widespread uptake of network technology during the 1990s has brought about a need for all professions to consider a disciplined approach to information; a need to avoid the feeling of 'information overload'. The approach described above proffers a framework for research which embraces resources in all formats, and which maps out a method for identifying and selecting relevant information, within a manageable structure. Even so, the task of finding resources for research can be time consuming, and demands skill on behalf of the user. Despite the ease of graphical search screens, the usual skills of subject knowledge and information science still help.

References

1. Gilster, P. *Digital literacy*. New York: Wiley, 1997, 181.

2. Gibson, C. and Meade, J. Internet resource discovery: the role of critical thinking. In: Leonhardt, T.W., ed. *'LOEX' of the West: teaching and learning in a climate of constant change*. Greenwich CT: JAI Press, 1996, 137–146.

3. OMNI *Guide to OMNI evaluation criteria*. 1999. http://www.omni.ac.uk/agec/evalguid.html (visited 10 January 1999).

4. Wexler, P. *Information resources in toxicology*. 2nd ed. New York: Elsevier, 1988.

5. Grogan, D. *Science and technology: an introduction to the literature*. London: Bingley, 1970.

6. Hill, M., McIlwaine, I. and Williamson, N. Series Editors' Foreword. In: Wyatt, H.V., ed. *Information sources in the life sciences*. 4th ed. London: Bowker-Saur, 1997.

7. NRCST *A directory of information resources in the United States: general toxicology*. Washington DC: National Referral Center for Science and Technology, Library of Congress, 1969.

8. Sullivan, D. www.searchenginewatch.com (visited 26 January 1999).

9. Anagnostelis, B. Thinking critically about information on the Web. *VINE*, 104(21), 1997, 27.

10. Tate, M. and Alexander, J. Reaching critical evaluation skills for World Wide Web resources. *Computers in Libraries*, November/December 1996, 49–55.

Notes

1. 'Search' implies querying data resources using search terms representing the toxicology (or sub-topic) concept, plus additional terms as necessary to define the kind/s of resource being sought, e.g. information resource list/listing/links /web.

2. Bibliographic databases may include:

- Subject specific: Toxline, Medline, Embase, Biosis, Life Science Collection, Scisearch.
- Library/information: LISA, Social Scisearch.
- 'General': Books in print, Dissertation Abstracts, NTIS.

3. Internet search tools based on those described as 'major' by Sullivan [8], which have different underlying indexes. Additional search tools, such as those that identify newsgroups and mailing lists, and those that are particularly good for identifying toxicology-related material, to be used where appropriate.

4. Level specific resources to include guides relevant to the type of literature being sought, e.g. Database/CD-Rom catalogues, library catalogues, lists of societies or journals.

Task analysis of the webmaster: results of an empirical study

Pieter Van der Walt
Altron Systems, Johannesburg, South Africa
pvdwalt@ams.altron.co.za

Pieter Van Brakel
Department of Information Studies, Rand Afrikaans University, Johannesburg, South Africa
pavb@lw.rau.ac.za

The functions or tasks of the typical webmaster appear to be extremely diverse in nature. These can vary from basic web page design to gathering website statistics or even to establishing the web-based information needs of the employees of a specific organisation. Although a number of mainly electronic documents were written on the functions of the webmaster, not many studies were undertaken to empirically verify these. This article discusses the results from an international survey in which the webmasters or managers of web teams were involved from 100 of the Global 500 list of organisations. An analysis of the information provided by the respondents indicated that one person (the traditional webmaster) can no longer be responsible for all the responsibilities of developing and/or maintaining the information flow within an organisation's website. The diversity of tasks that were identified by the respondents were analysed and the results brought together in a number of figures and tables.

Introduction

A comprehensive literature survey on the work-related responsibilities of the webmaster was published just over two years ago [1]. In the light of the growing strategic importance of a well-designed website to any enterprise, it was decided to launch an international follow-up research project in an effort to investigate how organisations have been addressing and developing the various tasks of the webmaster – now a relatively established career option. Academics responsible for offering training programmes to prepare graduates for this specialised career group should take notice of the results of this empirical study.

As the above-mentioned article discussed the term 'webmaster' at some length, the various definitions provided in the literature will not be discussed here again. Suffice to cite Richmond's inclusive description as a working definition for this article [2]:

> '... person who manages a web; mediator between web authors and system administrator – ensures that applicable standards such as HTML validity and link liveliness are met; optimises the web architecture for navigability; takes editorial responsibility for the content, quality and style of the site; finds, creates and installs tools to create web content; check consistency, develops and enforces the house style; liases with graphic artists; provides first level user support.'

From this definition it is clear that the typical webmaster is entrusted with a wide variety of different tasks during his or her day-to-day activities. Furthermore, the boundaries between these different roles are not yet clearly defined, as the webmaster is expected to perform different and sometimes unrelated functions such as authoring, systems administration or graphic design. What makes matters even more diverse, is that in a smaller organisation all these tasks might be performed by a single person, compared to a larger organisation where each role might be performed by several members of a web team [3].

Research design

To determine the specific responsibilities and tasks associated with the position of the webmaster, a questionnaire was compiled and posted on the Web (see Appendix A). The objective of the questionnaire was to 'interview' current webmasters working in a broad spectrum of the private enterprise, in an effort to provide the authors with different viewpoints of the position and the role of the webmaster within the organisation, as well as identifying the requirements necessary for a webmaster to successfully complete his/her various responsibilities.

The questionnaire consisted of the following sections:

- biographical detail;
- job title;
- job description – tasks;
- job specification – human skills;
- any other comments or suggestions.

Obtaining respondents

For the validity of the data it was necessary to reach as wide a spectrum of webmasters as possible. The Web itself was utilised to do exactly this. The respondents were selected by dividing the *Fortune Magazine's* Global 500 list of organisations (www.pathfinder.com/fortune/global500/500list.html) into five equal sections, that is organisations listed as 1 to 100, 101 to 200, 201 to 300, 301 to 400 and 401 to 500. By randomly selecting 20 organisations from each section a shortlist of 100 organisations was compiled. The complete list of the selected 100 organisations is available in Appendix B, along with their respective website addresses.

It should be noted that the Global 500 list had been chosen instead of the Fortune 500 list. The Fortune 500 list mainly covers organisations located in the United States of America, whereas the Global 500 list includes organisations from a global perspective and furthermore publishes the 500 largest organisations in the world [4]. Also, these organisations were not limited to a specific industry or geographical location and therefore could theoretically provide a very broad and international view of the position and tasks of the webmaster.

The webmasters responsible for the websites of these 100 organisations were contacted via electronic mail and individually invited to complete the web-based questionnaire. In order to test the functionality and content of the questionniare, five South African webmasters were requested to complete a preliminary release of the questionnaire. Respondents were given two weeks to submit their answers.

Analysis of results

At the end of the two-week period (that is, November 1998) only 28 of the 100 respondents had completed the questionnaire. A follow-up request was emailed and an additional 21 respondents responded. Some respondents indicated that it was not their organisations' policy to take part in surveys like this. In these cases other organisations were selected. Additionally, two respondents indicated that their websites were outsourced to external companies and therefore did not employ a webmaster. These organisations have been included in the results. Finally, after a third reminder a total of 63 respondents out of the possible 100 took part in the survey.

The following paragraphs deal with the data received from the questionnaire.

Biographical detail

Age

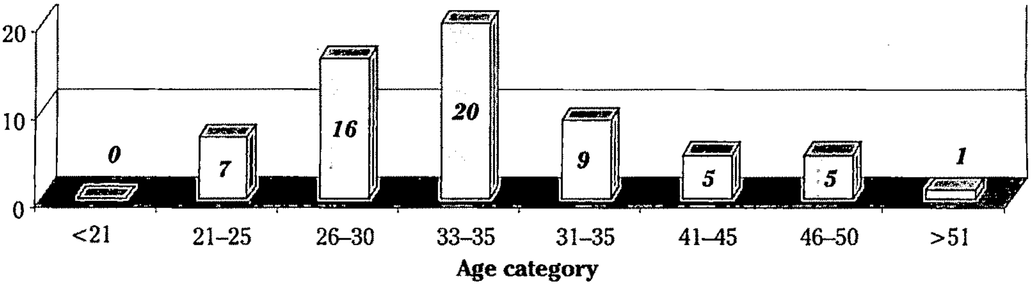
The age group was broken down in categories of 5 years, ranging from less than 21 years of age, incrementing with 5 years, up to over 51 years old. From the data provided by respondents, a webmaster age group distribution graph could be compiled (see Figure 1).

The majority of webmasters were aged between 31 and 35. The second largest grouping was even younger, that is, between 26 and 30.

The relatively young age of the webmasters goes hand in hand with the very brief existence of the Internet and World Wide Web. It can also be argued that it is thus the playground of a younger generation involved

P 9644

Figure 1. Webmaster age distribution



with the technological development surrounding the Web. Of the 63 respondents, 45 were male (18 female) indicating that this career is currently dominated by males.

Job title

Job designations ranged from web related titles (such as webmaster, home pages manager, web editor, global communications webmaster, website operations manager), to more general and broader IT-based titles (such as systems programmer, assistant manager, network project analyst and IT manager).

Regarding the department or division where the webmaster was operating, it was noticeable that there were two broad types of department or division, namely, a technical information-systems division and a marketing or corporate communications sector. Although these did not all carry the same names, their designations indicated a specific role within the organisation. Divisional or departmental names ranged from MIS (Management Information Systems), IS (Information Systems), Corporate Relations, IT (Information Technology), Customer Services, Internet Technology, Creative Multimedia Services and Corporate Services and Marketing.

If an organisation's website forms the focal point of its electronic presence, it would be the responsibility of the marketing or corporate relations division to ensure that the desired image of the organisation is portrayed to the Internet community. There are, however, skills of a more technical nature, such as programming and system administration, that would traditionally fall under the IT, MIS or IS divisions. Both these areas are of importance to an organisation, as it is the marketing division that indicates how they would like to portray the organisation's image on the WWW, whilst the information technology division provides the technical expertise and skills to establish the organisation's web presence.

Primary activity of employers

The organisations approached to participate in the survey were not limited to a certain industry sector, but ranged between areas such as manufacturing, software engineering, equipment engineering, information systems, document handling, consulting, electronics, insurance, telecommunications, aerospace, petroleum fuels, motor vehicle industry, software development and accounting.

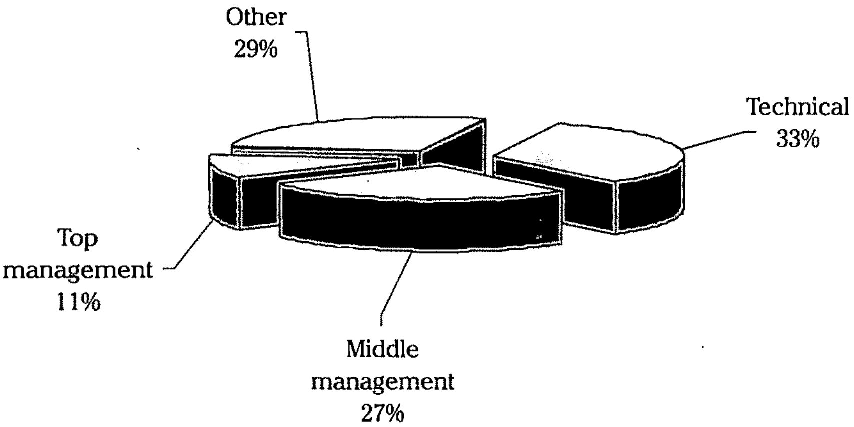
Length and type of employment

The average time of employment of the respondents as a webmaster was as brief as one year and eleven months (1.92 years)! The longest term of employment as a webmaster was five years. Of the 63 respondents, 60 were employed full-time and permanently, one part-time and two on a contractual basis as webmasters. As already stated, some respondents indicated that their WWW development was outsourced or contracted to outside companies. It became clear, however, that the majority of organisations have already realised the need for a webmaster or a webmaster team.

Position within the organisation

Findings with regard to the level of employment of a webmaster within the organisation are depicted in Figure 2. A large proportion of

Figure 2. Employment category



respondents were employed at a technical (or IT) level (in this case, 21 respondents or 33%). On the other hand, 17 respondents (27%) had middle management positions and 18 (29%) were undefined or mentioned in the category 'other', as defined within the questionnaire. These 'other' positions included posts such as individual consultants. In some cases, the webmaster was not employed full-time. An important finding was that 11% of the respondents held positions within the top management structure of their organisations.

It is a well-known fact that, in the 'early' days of web development, webmasters were responsible for technical tasks such as HTML markup, programming and web page design. The relatively high percentages depicted in Figure 2 (38% of the respondents were working in the middle and top management levels) is an indication of an important shift in responsibilities:

- It is a movement away from the more 'technical' tasks to an involvement in the managerial responsibilities of an enterprise.
- It can therefore be argued that the webmaster has become an integral part of a business and marketing venture of the average organisation, and because of that has become vital to the entire business strategy of an organisation.
- Also, that websites are certainly becoming mission-critical in the sense that they are seen as a business tool. The term 'mission-critical' implies that organisations can no longer afford not to have a presence on the Web.

Whereas 38% of the respondents hold a position in management (that is, combining the top and middle management sectors) a clear correlation between the position of the webmaster and the size of the organisation could not be established through the survey. However, the age distribution indicated a normal tendency, that is, the more senior the position, the older the person holding that position, in this case 31 to 35 and 36 to 40 respectively.

The web team

Sixty-nine per cent of the respondents indicated that they are assigned the responsibility of managing a team of staff members. The remaining respondents operated without the help of a team. In one instance, an organisation specialising in telecommunications technology indicated that their website was outsourced to an external company for development and upkeep. Correspondence received via this company's site was taken care of by the corporate communications division. This organisation was in the 400 to 500 bracket on the Fortune 500 ranking.

Amongst the organisations that did have a website team, the average size of the group was 13.7 members. However, it should be noted that two organisations indicated that they respectively had 50 and 100 members devoted to their sites! This would therefore influence this average substantially. In most cases organisations indicated that there was only one webmaster employed, and the smallest number of team members (including the webmaster) was two. However, there were also teams consisting of five, ten and twelve members, which included HTML programmers, graphic artists and networking specialists.

It seems as if developing and maintaining a website single-handed is no longer possible. It is therefore not unexpected that nearly 70% (69%, to be exact) of the participant webmasters manage a team where each individual specialises in a certain area.

As organisations start to see the importance of having an online presence through their websites, more strain is being put on the smaller website teams. These teams would soon find it necessary to start to expand through the addition of staff specialising in the various areas of site design and maintenance. This was especially evident from the results of the study, for example in cases where organisations specialised in entertainment and software development, as their websites have become an additional and versatile channel in the distribution of information.

Job description

The job description section of the survey addressed the different tasks assigned to the webmaster. Respondents were asked to indicate which of the tasks they were responsible for, specifically as head of the webmaster team, and alternatively, to indicate which tasks were allocated or delegated to other members of the team.

The results are presented here in the same order as were the questions in the questionnaire. Column A reflects the webmaster's responsibilities when managing a whole team, while Column B indicates a specific task or responsibility that applies to the other members of the team. The respondents, that is webmasters, were asked to complete Column B if a web team was present (that is, one or more additional members in the group).

Managerial tasks

In all cases, respondents indicated that, in managing a website development team or department, project management and information management played an important part in their job descriptions (see Table 1).

As can be expected, the A category (head of the section or department) got the highest score. More specifically, project management takes up the major part of the management task. Following this is the management of the actual website content. Team members also perform certain managerial tasks, as were delegated to them from time to time.

Editorial content creation and capturing

Surprisingly, it was found that webmasters are not primarily responsible for creating or writing the content that goes onto a website. In fact, the split was even, with both at 59%, as the trend is rather to appoint staff or ask certain employees of the specific organisation to create content (for example, to write about their section or department, purpose or any other topic that could be used to enhance the purpose of a certain website or web document) (see Table 2). It is only through this co-operation of the different divisions and sections that the webmaster could accurately communicate information regarding the organisation, in all its different forms, to the visitor of a site.

Graphic design and layout tasks

The results in Table 3 prove the importance of the graphic designer; some respondents indicated that a person with graphic design skills 'is crucial' to the success of the web team and, ultimately, the presentation of information on the organisation's website. They (the results) also indicated that the webmaster is no longer directly responsible for this task. In most cases the design and layout of the appearance or representation of the website was the responsibility of team members specialising in graphic design or page layout (desktop publishing or interface design) (see Category B in Table 3). The graphical appeal of a website plays a very important role in the presentation of a product or information. Also, being a relatively

Table 1. The management tasks of the webmaster

Management		**A**		**B**
Managing a website team or department	42	67%	14	23%
Managing projects	56	89%	16	26%
Managing website content	44	70%	26	41%

Table 2. Editorial tasks

Editorial		**A**		**B**
Content creation	37	59%	37	59%
Copy-writing and capturing information to be included within the website	33	52%	42	67%

Table 3. Design tasks

Design		**A**		**B**
Graphic design	28	44%	42	67%
Page layout (desktop publishing/interface design)	37	43%	35	59%
Multimedia design (audio and video integration)	26	41%	33	52%

laborious task, the design of graphics and/or multimedia effects is therefore in general left to other members of the web team. Most likely, this would be a dedicated graphic design artist or desktop publisher who specialises in the creation of graphics and layout of pages. This was confirmed by 67% of the respondents, indicating that the graphic design task was left to other members of the team (presumably a graphic design artist); 59% and 52% indicated that the team members are largely responsible for page layout and multimedia respectively.

Marketing and promotional tasks

Marketing and promotion implies planned actions to make the site known, thus increasing its traffic. The survey indicated that marketing is largely a group activity, although the webmaster also participated in the marketing of the website by, for example, following up on the accuracy of entries into different search engines. The actual entry of keywords or addresses into a search engine can initially be seen as a tedious process as the necessary keywords that best describe the website need to be identified. The webmaster's role was also to identify suitable websites where 'banner advertising' could be implemented to further promote the organisation's website. An emerging trend of these search engines is to provide related information in the context of the end-user's search. For example, when an end-user searches for 'telephones', the results returned to the end-user will also contain promotional material from companies within the telecommunications field. The aim is to launch more market-focused advertising and to make the end-user aware of related products or services within the context of the specific keyword search.

From the search results (see Table 4) it is also clear that the process of finding suitable advertising locations should be done by a person who knows the organisation and the search behaviour of the potential users – in this case the webmaster. Comments from individual respondents indicated a number of other initiatives to market the website, for example to ensure that the organisation's website address appears on all types of media used within and without the organisation (including letter heads, fax transmissions, business cards, general stationery and advertisements). Furthermore, as mentioned by some respondents, the webmaster can, in conjunction with the organisation's corporate

Table 4. Marketing tasks

Marketing		**A**		**B**
Website promotion (adding search engines' entries)	35	56%	33	52%
Website promotion via other media	28	44%	42	67%
Setting up and/or creating banner advertising	26	41%	35	56%

Table 5. System-related tasks

System tasks		**A**		**B**
System recommendation	40	63%	23	37%
System implementation	33	52%	30	48%
System administration	35	56%	33	52%
Log file analysis and reporting	37	59%	30	48%
Database administration	30	48%	35	56%
Website maintenance (fine tuning, monitoring)	44	70%	28	48%
Website security	33	52%	30	48%

communications or marketing division, initiate certain promotions for products on the organisation's website.

System administration tasks

System development and maintenance, be it hardware or software, forms the basis that drives the website. From the data received on the more specific questions asked (see Table 5) it would seem that the webmaster, as head of the team, scored substantially higher for tasks such as 'system recommendation' and 'website maintenance'. This type of finding could be expected, as assignments that have a direct impact on the continuity of the site should reside with the leader of the team, that is system development and maintenance. 'Log file analysis and reporting' can also be included in this argument, that is, analysing statistics and reporting to management.

Programming tasks

No longer are websites a static representation of information created solely from HTML. Browsers have become more complex and now allow for the interpretation of scripting and applications embedded within HTML. On the server side, the http daemon (httpd) has also become more complex. Instructions or scripts can now be executed on the server before returning the results to the end-user. These developments allow the end-user to carry out searches, set-up user profiles and carry out a variety of tasks through a web page. Organisations are also implementing browser technology to act as an interface into their legacy and database systems, or increasingly integrating their database systems with their Internet as well as intranet websites. Therefore, applications on the website are becoming more reliant on the programming skills of the webmaster and team.

Table 6. Programming tasks

Programming		**A**		**B**
Script writing (e.g. CGI, PERL, Java, JScript, ISAPI)	37	59%	28	44%
VRML	16	26%	28	44%
Tool or application development	28	44%	35	56%
Database integration	35	56%	30	48%

All this interactivity and integration with back-end systems requires the writing of scripts or special programs. Although web creation software is becoming easier to use and have multiple What-You-See-Is-What-You-Get (WYSIWYG) interfaces to WWW pages, there are still some finer-tuning programming issues that need to be addressed by the webmaster.

As indicated by the results in Table 6, a variety of programming tasks need to be addressed in this requires. Although 'script writing' was rated 59% the responsibility of the webmaster, this and the other results in the table do not indicate any prominent differences between the respective responsibilities of the webmaster and team. The fact of the matter is that today the Web provides more and more interactive communication possibilities. This was not previously possible using conventional HTML. Script writing and technological developments, specifically aimed at the WWW, have made this possible. From the results it would seem that programming is more of a team effort than the responsibility of one individual.

Very few respondents rated VRML skills (Virtual Reality Modelling Language) as an important programming task. This low figure is echoed by Parisi, Vice-president of the VRML Consortium, which can be contributed to lingering fears from the developer community, where most developers do not have the experience of creating VRML content [5]. Also, according to Parisi, before 1997 the use of a special browser to read or interpret the language hindered the use of VRML and associated applications browsers. It seems, however, that this figure will increase with the inclusion of VRML technology in the new generation of web browsers.

Support and training

As expected, the webmaster was less involved with training and supporting end-users and more involved with the training of specific web staff members (63%). Team members, on the contrary, would be more able to support or train the end-user. The webmaster, due to experience, knows the organisation's structure and culture and, through this, acts as a mentor to other WWW team members in terms of education. Furthermore, the webmaster in his or her capacity as researcher and developer is in the logical position to transfer details of new technology to other members of the WWW team.

Research and development

To keep abreast with new technological developments and developments amongst the competition, it is necessary, in any environment, to research new technological developments and technology trends. Three items were identified to represent the need for R&D: system technology, website technology and technology deployment. Results (67%, 52% and 67%) indicated that the webmaster was primarily responsible for undertaking these tasks (see Table 8). This could mean that not many organisations could as yet afford the luxury of appointing a person specifically for R&D purposes. The webmaster, with his/her more senior involvement with the site, is the person to initiate such activities. However, no correlation could be found between the number of team members and the position of the organisation in the Fortune 500 ranking. The fact that nearly all respondents indicated that research and development is an important task confirms that the webmaster and his/her team need to have a thorough understanding of the organ-

Table 7. *Support and training tasks*

Support and training		**A**		**B**
End-user support	26	41%	35	56%
End-user training	19	30%	42	67%
WWW staff training	40	63%	24	33%

Table 8. Research and development tasks

Research and development		**A**		**B**
Website technology research	42	67%	26	41%
System development	33	52%	26	41%
New technology deployment	42	67%	21	33%

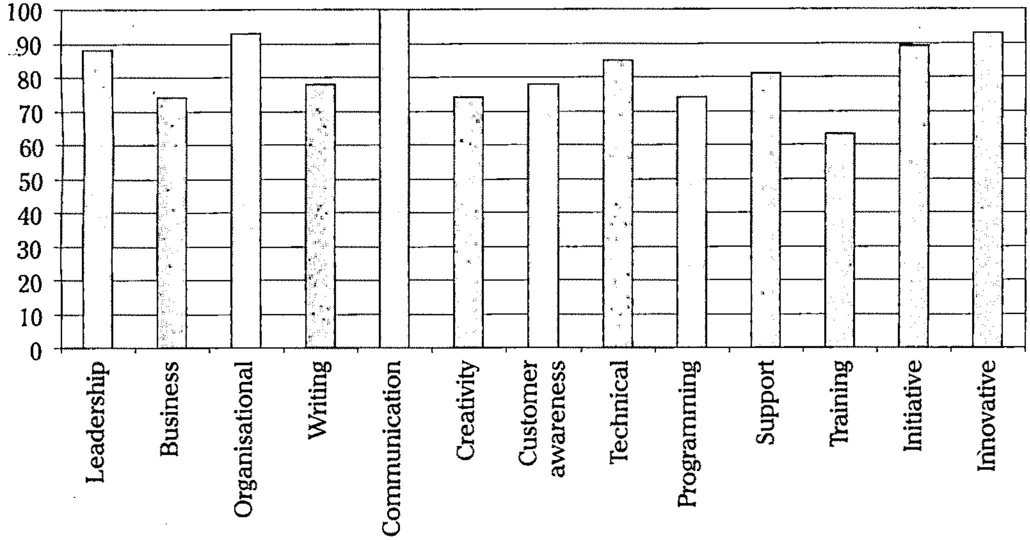
isation, and keep up with and recognise the significance of emerging technologies and business models. From the participants' contributions, it seems clear that the webmaster should be able to inform and educate the rest of the organisation and to demonstrate the business benefit of new technologies. This is especially evident in organisations that rely on technological innovation to enhance their products and services.

Personal comments by respondents

The questionnaire (see Appendix A and specifically 3.9) enquired about any additional tasks that respondents would like to include and which did not appear in the questionnaire. The following personal contributions were recorded:

- Communication: Two organisations, respectively involved in software development and document management, responded that answering electronic mail from end-users (sent via the website) formed a crucial task. Communicating with end-users was thus an important function of the web team. In the case of the software development organisation, the website formed a reference source to many end-users, especially for technical information. However, if the end-users' needs were not met, they could send requests or opinions to the webmaster who could then route these requests or opinions to the correct divisions within the organisation.
- Website policy: A large telecommunications organisation indicated that policy enforcement or policy setting was an important task of the webmaster, to ensure continuity of website and content. Within a large organisation, where numerous divisions contribute to the organisation's central website, it is therefore necessary to ensure that a standardised look and feel of the website was enforced and that all electronic documents adhered to certain standards. On the other hand, there are also numerous organisations within a large corporate entity, and each organisation has its own website. Therefore, certain corporate standards have to be enforced if that is the policy of the corporate entity.
- A large technology organisation ranked within the top 200 bracket indicated that strategic decision-making is another essential responsibility of the webmaster within the larger organisation. The latter consists typically of business units or smaller functional organisations. The webmaster acted as the convenor of all the different participating parties within the organisation, with regard to the organisation's international WWW presence. The webmaster obtained input from the different divisions and provided guidance in the formation of the organisation's online presence.
- Account management of various content owners was seen as a key responsibility of the web team members, especially in large organisations with numerous products represented on the organisation's website. As mentioned previously, the webmaster obtains input from various organisational units and puts this information together into a single website, portraying the overall image of the organisation.
- The personalization of websites and information resources has been identified as the webmaster's task by a large software development organisation. To provide personalised information to the end-user requires the application of push technology or web casting, which gives the end-user the choice of receiving information that is specific to his or her needs.

Figure 3. Required skills of the webmaster (both A and B groups)

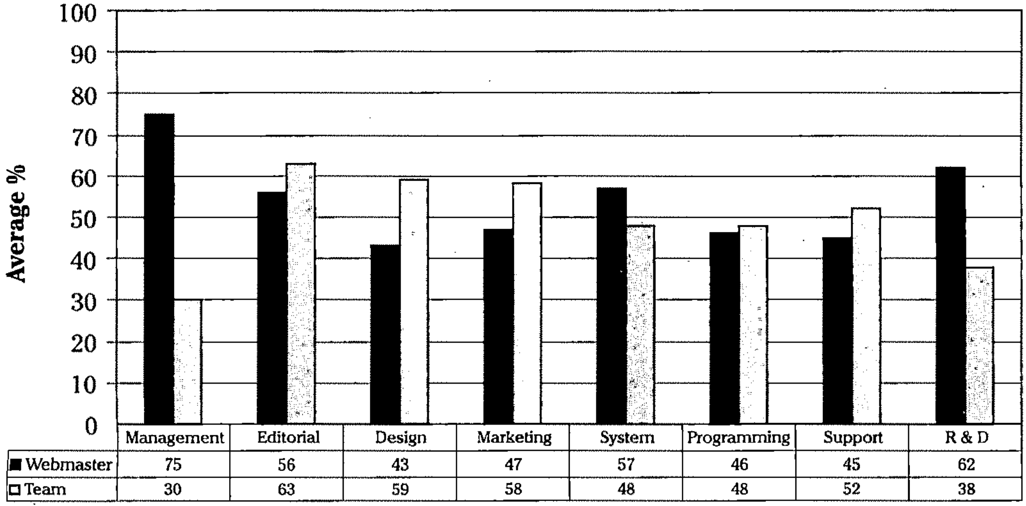


Skills needed in a web office

This section of the investigation (see Appendix A, Question 4) requested the respondents to indicate which job skills he or she feels are necessary to be able to perform effectively in a web team environment. This was to supplement the responses on the various responsibilities of the webmaster and team (see above), and to establish which unique training needs exist for this particular career group. Again, respondents were asked to select the options that apply to themselves as well as to the website team.

Figure 3 represents the total percentages

Figure 4. Division of task categories between A and B



	Management	Editorial	Design	Marketing	System	Programming	Support	R & D
■ Webmaster	75	56	43	47	57	46	45	62
□ Team	30	63	59	58	48	48	52	38

indicated by the respondents, indicating all the job skills identified during the survey. It was an important finding that except for 'Training' (which came in just under 65%), the vast majority of skills were mentioned by more than 70% of the respondents.

When comparing the various broader categories (see Appendix 4, paragraph 4) and according to the same A (webmaster) and B (the team) categories, the following figure could be drawn (Figure 4).

Certain categories, especially management, system, and research and development, appeared to be, according to the respondents, the primary tasks of the webmaster as manager of the website. The remaining tasks, on the other hand, (editorial, design, marketing, programming, support, and training), tend to be associated with the website team. However, programming responsibilities seem to be closely divided as a responsibility of both the webmaster and the web team.

The following paragraphs discuss these findings in more detail (see Figure 3).

Management

Possessing organisational skills appears to be a logical but essential qualification for a good webmaster. Being involved with numerous projects requires that a person has good organisational skills, together with a high level of leadership skills. Business skills are also needed to provide good management through knowledge of the business environment and a background knowledge of the organisation. The following ratings were acquired in the survey:

- Leadership: 88%
- Business skills: 74%
- Organisational skills: 93%.

The importance addressed to these skills endorses the respondents' identification of the tasks as discussed in the previous section.

Editorial

In conjunction with content creation and data capturing, writing and communication skills were deemed necessary to fulfil the task of representing information on the WWW, and communicating directly to the end-user. Seventy-eight per cent of the respondents indicated that good writing skills were a requirement for the webmaster. Together with this, all respondents (100%) indicated that communication skills were a necessity for the webmaster. These two skills in themselves are important, not just for representing information correctly and accurately on the organisation's website, but also to insure that the information is correctly targeted at the intended market or end-user. Organisations that specialise in producing a product range from a specific market need to make sure that the website reflects the market for which that product is intended. This should be seen in conjunction with the marketing and design of the website.

Design

Seventy-four per cent of the respondents indicated that creativity was a required skill for the webmaster. In the web environment the webmaster can bring different formats of media together into a single entity, by combining text, graphics, sound and movement on one web page and thereby give a totally new meaning to a document, product or service compared to what is available on conventional media. The use of multimedia in all its forms, through images, animation, sound and video, ensures that websites are more than just informative but also entertaining to the end-user. This might include a combination of multimedia effects to promote a website or product via the electronic medium of the Web.

The relatively high indication from respondents (72%) indicates that creativity and originality in design play a very important role in the visual appeal of a website. It is not only the design of certain graphical aspects, but it also includes the spatial layout of visual elements on the website. These visual elements can consist of text, graphics and multimedia. The ability to present information in a number of ways on the WWW lends itself to being used as a canvas, where the webmaster can experiment with his or her creativity by using all forms of multimedia. This also enables the webmaster to become part of the organisation's strategy in promoting products or services through a diverse range of communication mediums.

Marketing (or customer awareness)
Being aware of the end-user's needs can help in the development and design of the website. Seventy-four per cent of respondents indicated that promotion and presentation of products of information is of particular importance. This is also true in the case of targeting a specific market of end-users. Websites should always be designed with the end-user in mind.

System (technical skills)
It is not surprising that 85% of the respondents indicated that system management and administration were important skills. In order to ensure the successful running and functioning of a particular website, it was required of the webmaster to have the necessary technical know-how to understand the underlying hardware and software forming the core component of the website.

The technical nature of the Web requires a person with technical skills to maintain such an infrastructure and ensure that the organisation's website is available to the end-user for 24 hours of the day. To many organisations the Web is becoming a crucial business tool and therefore requires a person or persons who are technically inclined and capable of effectively maintaining the organisation's website and underlying infrastructure.

Programming skills
Seventy-four per cent of the respondents indicated that programming skills were important. This is not surprising seeing that websites are becoming more and more interactive in terms of accessing databases via scripting or CGI. HTML is not a very flexible language in the sense that it does not primarily support high-level programming characteristics such as C++ or Delphi. However, HTML can co-exist with applications that have been written in these high level languages as well as being able to support variants of scripting languages embedded within the HTML.

This implies that the webmaster and team should, in addition, be skilled in a programming language such as Visual Basic, VBScript, Java, JavaScript or C++. This enables HTML to be taken further than just publishing static information. It also enables the end-user to interact in a far more flexible way with the website through applications and scripts than was previously possible with generic HTML.

Support (communication and training)
To communicate clearly and effectively with other people, be it team members or end-users, it is important to ensure that the correct information is presented to the correct people. Eighty-one per cent of the respondents rated communication skills as part of the required skills for a good webmaster. Training skills performed substantially lower at 63%, indicating a relatively low importance with regard to in-house or in-service training.

R&D (initiative and innovation)
The last category falls under research and development and more specifically, skills such as the ability to use initiative and have an innovative mind. The web is evolving constantly, and new innovative ways are being sought to present and promote products and information on the WWW. Webmasters with enough initiative and innovation would make sure that new technology and ideas could be implemented. This is being proved by the high scores allocated by the respondents for these two elements: 'Initiative': 89% and 'Innovative mind': 93%.

Other skills and additional comments
As in the previous section, respondents were asked for additional comments with regard to the skills required in a webmaster's office. The following items were specifically listed by some of the respondents:
- patience;
- teamwork capabilities;
- dedication to complete tasks;
- perseverance;
- motivation;
- independence;
- leadership qualities.

Further comments were made by the respondents in the 'general' column – these are listed below, as they represent the personal experience of practising webmasters.

- In any organisation it is the master who is responsible for monitoring the corporate web pages describing these divisions, for example, with regard to branding and content inaccuracies. Knowledge of the organisation's structure and strategy is an essential element, of course. It is argued that the webmaster is in the ideal position to develop a strategy to monitor any changes.
- The webmaster should be self-motivated to enable him/her to stay abreast of the latest technologies. This aspect goes hand-in-hand with the research and development skills required of the webmaster as discussed previously. In order for the webmaster to stay ahead he or she must constantly read about new technologies and investigate the possibilities to incorporate them into the organisation's website.
- WWW evangelism is a critical skill to ensure funding and also to elicit participation from different players to work together for common website goals. In order for the organisation's website to be successful, the webmaster needs to know what the end-user's needs are with regard to online applications. Furthermore, the webmaster needs to propagate the facilities provided on the organisation's website to the broader end-user community. This can be accomplished through seminars, newsletters and the normal marketing techniques.
- Together with project management, the webmaster needs to be able to prioritise tasks and to complete them efficiently and on time.
- The webmaster should pay attention to detail. In many respects, the organisation's intranet has become the central source of internal information, such as the organisation's procedures, human resources and various other applications. On the other hand, the organisation's public website projects the image of the organisation, its products and services to the public end-user. The webmaster needs to ensure that the information contained within the website is correct and accurate. Furthermore, the webmaster also needs to make sure that the integrity of the website, the style of the site, navigational aids and any other aspects are implemented throughout.
- Educational background is less important than experience; '...the educational system lags behind currently available technology on the Web...', one of the respondents remarked. Institutions that provide education or training find it difficult to keep up with developments due to the implementation and development of new technologies within the online field. Furthermore, problems are due to the unprecedented rate at which change is taking place. In many cases, courses become obsolete with the development and ultimately the replacement of older technologies by new ones.
- From a managerial point of view, the corporate webmaster does not need to be a graphic designer, a systems analyst or even a programmer. However, he or she has to have a basic understanding of the principles guiding each of the areas covered by webmastering.
- The webmaster should be able to communicate with executives or management in terms that they understand or can relate to in the business environment. Another critical aspect is the ability to translate the jargon associated with the WWW or 'web speak' into terms that executives or management can understand and embrace. For example, various terms describe certain processes regarding websites. These terms mean something specific to programmers or developers and are of lesser importance to executives or management. To translate these into understandable laymen terms requires good communicative skills.
- In cases where there is a large organisation providing information on the Web, the webmaster fulfils the role of overseer in monitoring website content, branding inaccuracies, such as correct logo or trademark usage, and to provide guidelines and strategies regarding the website presence.
- Finally, in cases where a single person is responsible or dedicated to an organisation's website, the webmaster acts as a liaison between organisations being outsourced and the divisions within the

organisation to provide the necessary information for the website.

Conclusion

From the input and additional comments received from the Fortune Global 500 respondents, it could be established that multidisciplinary skills are required in any webmaster's office. Also that the webmaster and team play an important role in establishing and ultimately maintaining an organisation's online presence on the World Wide Web. However, by looking at the different aspects that make up a website, such as data capturing, graphic design, development, presentation, multimedia, legal issues, hardware and software, the role of the webmaster becomes more and more diverse. It seems as if it would be more beneficial for large organisations to establish a web unit or division, and thus form a central hub for the online or web presence of the organisation. This can be supplemented by establishing an internal web or Internet advisory committee (or even departmental working groups) where input is obtained from the different business units within the organisation.

As indicated by the results of this study, the role of the webmaster is relatively complex. It became clear that it is difficult to obtain a single person that has all the necessary skills. A basic understanding of the principles guiding each of these areas is, however, essential. In order to gather all these skills within a single unit, it became clear that different people should be employed to specialise in these different areas. The main responsibility left to today's webmaster is to provide the necessary guidance and management in running a well-functioned web office. An extremely important but relatively new assignment refers to the creation of the corporate intranet, with integrating and managing relevant strategic content into this environment.

Further research is needed on how the corporate intranet could and should be utilised for effective information management and knowledge management.

References

1. Van der Walt, P.W. and Van Brakel, P.A. The webmaster: a new player in the information centre's online team. *The Electronic Library*, 15(6), 1997, 447–454.

2. Richmond, A. *What is a webmaster?* 1996. http://www.stars.com/W3C/Provider/.

3. Sullivan-Trainor, M. *Webmaster strategies.* Foster City: IDG, 1996.

4. Fortune. The Fortune global 5 hundred – the global 500 list. *Fortune Magazine.* 1997. http://www.pathfinder.com/fortune/global500/500list.html

5. Dunlop, A. *Whatever happened to VRML?* 1998. http://www.internetworld.com/daily/trends/1998/01/3002-vrml.html

APPENDIX A

Questionnaire: task analysis of the webmaster

Please take a few minutes to fill in the requested information. Note that all correspondence will be treated confidentially!

1. Biographical Detail

First Name:		
Last Name:		
Age	< 20	☐
	21 – 25	☐
	26 – 30	☐
	31 – 35	☐
	36 – 40	☐
	41 – 45	☐
	46 – 50	☐
	51 >	☐
Gender	M/F	☐
Title	Mr/Mrs/Ms/Dr/Professor	☐
Email address		
Website address	http://__________	

2. Job Title

What is your job title?	☐
Name of the Company?	☐
Department/Division in the company	☐
What is the primary activity of your company?	☐
How long has Yours truly, been employed as a 'webmaster'?	☐

How are you employed?

Full time ☐ Part time ☐ Contract ☐

What is your current 'level' within the company?	Technical	☐
	Middle Manager	☐
	Top Manager	☐
	Other – specify	☐

Do you manage a website team?	Y/N	☐
If your answer was 'yes' in the previous question, how many members are there in this team?		☐
		☐
		☐

3. Job Description – Tasks

Which of the following aspects are you responsible for in relation to your specific task as a webmaster?
Select the options that apply to you as a manager of a website team and mark your selection in column A. If you feel that an option applies to another member of the website team, mark that option in Column B. Select all options that apply to you.

3.1. Management	**A**	**B**
Managing a website development team/department	☐	☐
Managing projects	☐	☐
Information management – managing website content	☐	☐
	☐	☐
3.2. Editorial		
Content creation	☐	☐
Copy writing and capturing information for the website	☐	☐

3.3. Design
Graphic design ☐ ☐
Page layout (desktop publishing) ☐ ☐
Multimedia design (audio and video integration) ☐ ☐

3.4. Marketing
Website promotion by adding entries into search engines ☐ ☐
Website promotion via other media ☐ ☐
Setting up and/or creating banner advertising ☐ ☐

3.5. System
System recommendation ☐ ☐
System implementation and configuration ☐ ☐
System administration ☐ ☐
Log file analysis and reporting ☐ ☐
Database administration ☐ ☐
Website maintenance – fine tuning, monitoring ☐ ☐
Website security ☐ ☐

3.6. Programming
Script writing/CGI (Perl, Java, JScript, etc.) ☐ ☐
VRML ☐ ☐
Tool development ☐ ☐

3.7. Support and training
End-user support ☐ ☐
End-user training ☐ ☐
Web staff training ☐ ☐

3.8. Research and development
Website technology research ☐ ☐
System development ☐ ☐
New technology deployment ☐ ☐

3.9. Are there any other tasks?

4. Job Specification – Human skills

Select the skills that you feel deemed necessary to fulfill the above-mentioned job description

Options selected apply to yourself and the rest of the website team.

4.1. Management
Leadership ☐
Business skills ☐
Organisational skills ☐

4.2. Editorial
Writing skills ☐
Communications ☐

4.3. Design
Creativity – creating the necessary visual appeal for a website ☐

4.4. Marketing
Customer awareness (aware of end-user requirements) ☐

4.5. System
Technical skills (knowledge of software and hardware) ☐

4.6. Programming
Skills to integrate web front-ends with databases, legacy systems via scripting/CGI ☐

4.7. Support and training
Communications skills ☐
Training skills ☐

4.8. Research and development
Initiative ☐
Innovative mind ☐

4.9. Are there any other skills that apply?

5. Please feel free to add any other comments or suggestions.

☐

Thank you once again for your co-operation. The final results of the investigation will be made available at this website towards the end of this year. You will be notified once the results are finalised.

Appendix B

List of Fortune Global 500 Companies

A complete list of the 100 of the 500 Fortune organisations selected to take part in this study and their respective website addresses. A complete list of the Fortune Global 500 organisations can be found at: http://www.pathfinder.com/fortune/global500/500list.html.

Rank	*Company*	*WWW address*
1	General Motors	http://www.gm.com/
4	Mitsubishi	http://www.mitsubishi.com/
10	Toyota Motor	http://www.toyota.com/
12	General Electric	http://www.gec.com/
15	IBM Corporation	http://www.ibm.com/
17	AT & T	http://www.att.com/
19	Mobil	http://www.mobil.com/
21	British Petroleum	http://www.bp.com/
33	Sony	http://www.sony.com/
36	Nestle	http://www.nestle.com/
40	Tomen Corporation	http://www.tomen.co.jp/
44	Texaco	http://www.texaco.com/
51	Philips Electronics	http://www.philips.com/
54	Fujitsu Limited	http://www.fujitsu.com/
55	Du Pont	http://www.dupont.com/
60	Hewlett Packard	http://www.hp.com/
65	Procter & Gamble	http://www.pg.com/
68	ABB Asea Brown Boveri	http://www.abb.com/
82	BASF	http://www.basf.com/
83	Bayer	http://www.bayer.com/
101	Motorola	http://www.motorola.com/
109	Hyundai	http://www.hyundai.com/

Rank	*Company*	*WWW address*
112	Lockheed Martin	http://www.lmco.com/
119	Kroger	http://www.kroger.com/
121	Merrill Lynch	http://www.ml.com/
130	British Telecom – BT	http://www.bt.com/
136	AB Volvo	http://www.volvo.com/
144	Boeing	http://www.boeing.com/
147	United Parcel Services	http://www.ups.com/
152	Johnson & Johnson	http://www.johnsonandjohnson.com/
155	GTE	http://www.gte.com/
161	Intel	http://www.intel.com/
162	BCE	http://bce.ca/bce/
180	Xerox	http://www.xerox.com/
191	Ericsson	http://www.ericsson.com/
192	Walt Disney	http://www.disney.com/
196	Coca-Cola	http://www.cocacola.com/
197	MCI Communication	http://www.mci.com/
198	Commercial Union	http://www.commercial-union.co.uk/
200	Compaq Computer	http://www.compaq.com/
203	Bridgestone	http://www.bridgestone.com/
205	Saint-Gobain	http://www.saint-gobain.com/
208	AMR	http://www.amrcorp.com/
216	LG International	http://www.lgeus.com/
218	Safeway	http://www.safeway.com/
222	AETNA	http://www.aetna.com/
225	Rhone-Poulenc	http://www.rhone-poulenc.com/
234	Fleming	http://www.fleming.com/
245	Eastman Kodak	http://www.kodak.com/
258	SNCF – French Railways	http://www.sncf.fr/
264	Bristol-Meyers Squib	http://www.bms.com/
266	Broken Hill Proprietary	http://www.bhp.com.au/
267	Ameritech	http://www.ameritech.com/
273	Bertelsmann AG	http://www.bertelsmann.de/
275	Digital Equipment	http://www.digital.com/
281	Coles Meyer	http://www.colesmyer.com.au/
283	Michelin	http://www.michelin.fr/
284	Rockwell International	http://www.rockwell.com/
297	SBC Communications	http://www.sbc.com/
298	Usinor	http://www.usinor.com/
302	Lufthansa	http://www.lufthansa.com/
310	Delhaize Le Lion	http://www.delhaize-le-lio.be/
313	NYNEX	http://www.nynex.com/
315	Sysco	http://www.sysco.com/
320	Enron	http://www.enron.com/
322	Norsk Hydro	http://www.hydro.com/
327	Goodyear Tire & Rubber	http://www.goodyear.com/
331	Glaxo Wellcome	http://www.glaxowellcome.co.uk/
333	Georgia-Pacific	http://www.gp.com/
346	Grand Metropolitan	http://www.grandmet-community.co.uk/
352	Marks & Spencer	http://www.marks-and-spencer.com/
365	Viacom	http://www.viacom.com/
377	Walgreens	http://www.walgreens.com/
378	Texas Instruments	http://www.ti.com/
390	UPM-Kymmene	http://www.upm-kymmene.com/
396	Mitsui Fudosan	http://www.ampark.co.jp/mfg/english/mitsui.html
397	Deere	http://www.deere.com/
399	Abbey National	http://www.abbeynational.co.uk/

Rank	***Company***	***WWW address***
404	Ikon Office Solutions	http://www.ikon.com/
408	Fluor	http://www.fluor.com/
410	Tenneco	http://www.tenneco.com/
413	Kawasaki	http://www.khi.co.jp/
421	Henkel	http://www.henkel.com/
429	Daiwa House Industry	http://www.daiwahouse.co.jp/
430	CSX	http://www.csx.com/
440	Waste Management	http://www.wastemanagement.com/
441	TRW	http://www.trw.com/
443	Federal Express	http://www.fedex.com/
448	Halifax	http://www.hxcorp.com/
454	British Aerospace	http://www.bae.co.uk/
456	United HealthCare	http://www.uhc.com/
457	Time Warner	http://www.pathfinder.com/
459	Union Pacific	http://www.up.com/
461	Bergen Brunswig	http://www.bergenbrunswig.com/
466	SUN Microsystems	http://www.sun.com/
469	Apple Computer	http://www.apple.com/
478	Cable & Wireless	http://www.cwplc.com/
483	Havas S.A.	http://www.havas.fr/
486	ASDA Group plc	http://www.asda.co.uk/
489	Principal Financial Group	http://www.principal.com/
490	Degussa	http://www.degussa.com/
496	Monsanto	http://www.monsanto.com/

Teaching online information retrieval to students of journalism

Charlotte Wien
Department of Journalism, University of Southern Denmark, Odense, Campusvej 55, 5230 Odense M, Denmark
Chw@journalism.sdu.dk

In the beginning of the 1990s it was decided to start educating journalists at two Danish universities. This was decided in the hope that such academic journalists would carry greater insight into that work. Therefore for the curricula of these students of journalism some traditional academic disciplines were introduced. One such academic discipline was information retrieval. It was clear, however, that the information retrieval course had to be designed specifically for the students of journalism, as the information needs of journalists differ from the information needs of more traditional academic disciplines. Thus, this article describes the work done in order to develop such a course for students of journalism. Firstly, it analyses the information needs of journalists on a theoretical basis. Secondly, an empirical study analyses which information retrieval resources are available to Danish journalists. Finally, it combines the theoretical and the empirical findings in arguing that it is necessary to provide students of journalism with a theoretical understanding of how online information retrieval works, and also practical experience with several information retrieval resources that they are supposed to use in their daily work.

Introduction

New knowledge is most often produced by combining information from different sources. To produce knowledge, therefore, requires information retrieval skills. An information retrieval skill is defined as the ability to find information in such a way that non-relevant data (noise) are excluded while relevant information is found. Today new information is mostly produced by researchers and journalists. The major difference between these two information providing groups is the amount of time available to them for producing new information. While the researcher might have three years or more to write a doctoral dissertation, journalists, on the other hand, have on average between three hours and one week to draw up an article. The researcher, therefore, might be satisfied with the retrieval of references, while the journalist usually will require full-text documents or facts.

Nicholas and Martin [1, p. 43] state that newspaper journalists:

> '. . . are very much in the forefront of the information technology revolution. Any changes in their information needs and information seeking behaviour could provide an early warning sign for other information communities,...'

This is one of the main reasons why this article will focus on newspaper journalists and their information needs. In terms of information retrieval journalists in general have two options: searching for information themselves or asking an intermediary (librarian) to do the search. Applying the latter option, they must be able to express their information need in such a way that the intermediary can develop a search profile from the information need statements.

Having retrieved some documents, the journalist or the intermediary has to decide on which of the retrieved documents are relevant to the information need. This is not as simple as it sounds. Some researchers [2, p. 50–53] within Library and Information

Science (LIS) consider the situation in which a user has an information need as an Anomalous State of Knowledge (ASK): the user does not really know what he[1] wants in order to satisfy his information need. The user has an information need *because* he does not feel that he knows all he needs to know about the subject. Many researchers of LIS think that the judgement of relevance necessitates knowledge about the subject of the information need. This has led some researchers to point out that '. . .it may be possible that the user is distinctly *unqualified* to judge relevance on the articles he has retrieved [4, p. 202–204]'. That information retrieval is a complicated process is reflected by the entire discussion of the concept of relevance and the discussion of user information need within the subject of LIS.

Bearing on the information needs of journalists the findings of Nicholas and Martin [1, p. 44] support that the problems discussed by researchers of LIS certainly apply to newspaper journalists. They found that:

> 'Journalists ... often provide vague and generalised descriptions of what they want, [to the intermediary, partly because . . .] vague subject specification mirrors the journalists' own uncertainty as to what they want . . .'

The only possible way to avoid such a situation seems to be teaching students of journalism the necessary information retrieval skills. Nobody has ever questioned the necessity of training the intermediaries. It takes several years to become a librarian. Only a few educational institutions, however, include student training in information retrieval. Basically, such training can be organised in three ways:

1. By teaching the students how to retrieve information from a series of information retrieval resources.
2. By teaching the students how to express their information need in such a precise manner that an intermediary can transform them into search statements and assist in the process of deciding which of the retrieved documents are relevant.
3. By combining the two approaches. That is, by teaching the students the basic theoretical problems involved in information retrieval regarding the expression of information need, and by applying the acquired theory to exercises in various information retrieval resources.

Which of the approaches to choose depends on the job situation of the students subsequent to graduation. Approach (1) is useful if it is known with certainty which information retrieval resources will be available to them when they graduate. Whilst approach (2) seems to be the most appropriate if the students can expect to have information retrieval support available to them for all information retrieval tasks. Finally, approach (3) should be used in the case of uncertainty regarding the conditions, in terms of online information retrieval support and resources, the students will face in their future job.

The subject of this article is the matter of choice between the three basic approaches mentioned in teaching information retrieval to students of journalism.

Background

Nicholas and Martin have carefully studied the information needs of journalists. They [1, p. 44] begin by dividing the user group 'journalists' into three subgroups, stating that their information needs differ dependent on the kind of newspaper they are working for. In particular journalists working for daily or Sunday newspapers do have the time necessary to research stories. This group of journalists, consequently, has the greatest need for information retrieval skills. Nicholas and Martin [1, p.44–45] found that the information need of this group can be divided into five broad categories:

1. fact checking;
2. current awareness;
3. researching;
4. obtaining context;
5. stimulus.

Of these categories, fact checking, researching and obtaining contexts are the only ones that a teaching program can include, because only these three presuppose systematic information retrieval behaviour. In the case of stimulus and current awareness, one might assume that browsing would be fully satisfactory.

Nicholas and Martin [1, p. 45–46] found, that there is no ideal label for the various types of information (conceptual, theoretical, historical, descriptive, statistical, or methodological) that journalists request, and that the journalists are generally capable of penetrating even the most complicated texts. However, the viewpoint or angle of a text may be of great importance to them, because they usually demand information written from the point of view of their own work.

The study of Nicholas and Martin shows that one of the main problems for journalists is the fact that they are bombarded with information. They see it as a major problem to cope with this excess supply of information. This explains why it seems difficult to convince journalists of the value of learning information retrieval. Recapitulating the point made in the beginning of this article that journalists lack the ability to state their information needs in a clear manner, and the findings of Nicholas and Martin, these findings support the view that the journalists are in need of a method for expressing information need in an exact manner. Such a method would simultaneously help the journalists to separate information into relevant and non-relevant categories.

In short, daily and Sunday newspaper journalists need information retrieval skills for fact checking, research and obtaining context. Their needs do not relate to any particular type of information. However, they prefer information that fits in with the point they are trying to make. And finally, one of the major problems for journalists is the lack of ability to express their information needs in a precise manner.

The choice between the three approaches mentioned above is conditioned by knowledge about the online information retrieval resources used by, or available to, journalists. The study of Nicholas and Martin [1] does not focus on this. But it seems obvious that FT Profile as well as wire services like Reuters and Associated Press are of great importance. Another related study by Nicholas and Williams [5, p. 7] shows that NEXIS, Dialog, Textline and Business Briefing are also widely used for information retrieval by journalists. And finally the importance of the Internet is growing at a tremendous speed.

Most educational institutions for journalists obviously aim at educating journalists in a national context. This very national context must be an important determinant in deciding on the information retrieval tools that the students of journalism are to be trained in using. Thus, for the purpose of teaching information retrieval to students of journalism it is necessary to focus on one particular country. For this article, Denmark was chosen. This was due to several factors. Firstly, because the country is rather small only a few institutions are educating journalists. Secondly, because the number of newspapers is limited. And thirdly, because the author is responsible for developing a course in information retrieval for students of journalism at the University of Southern Denmark.

Journalists in Denmark have for several decades been educated at one particular educational institution (Danmarks Journalisthøjskole) which is dedicated to educating journalists only. In the middle of the 1990s it was decided to break the monopoly of this institution by allowing two Danish universities (University of Southern Denmark and Roskilde University) to educate journalists also. The aim was, amongst others, to introduce some of the traditional academic disciplines into the curricula of students of journalism and hopefully to educate a new type of journalist – one who would be carrying greater insight.

One new academic discipline introduced at the University of Southern Denmark was 'Information Retrieval and Content Analysis'. It was clear from the outset that this course could not and should not be an exact copy of the courses given to students of more traditional academic disciplines. These students are mainly taught how to use reference databases, printed or online bibliographies within a specific scientific field. Given the time constraint of journalistic work and the fact that journalists might require information on practically every possible subject on earth, such course work would be a complete waste of time.

Another important aspect to be considered is the fact that no one knows for certain which information retrieval tools will be available to the students subsequent to graduation. Thus, there is no point in teaching them how to use tools available to journalists in 1999 when these tools might be outdated in 2001. New and much more effective ones might have emerged then. They have to be taught in such a way that their knowledge will not be outdated in two years' time.

A survey of the existing textbooks showed that there are basically two types:

- User manuals or tutorials for various information retrieval systems, or materials describing the content of selected information retrieval resources. Using such materials would only lead to an inductive method of teaching, and it would lead also to the choice of approach (1) mentioned above.
- Theoretical introductions to the scientific field of information retrieval in general. Using such material would lead to a deductive method of teaching and would imply the choice of approach (2) mentioned above. An example of such a textbook is the excellent one by Harter [6].

No available teaching material seemed to combine theoretical problems with practical exercises on selected information retrieval resources that would be relevant to journalists. If approach (3) was selected a new textbook therefore had to be written. However, selecting the basic approach to teaching the students of journalism information retrieval necessitated a study revealing which information retrieval resources and support were in fact available at Danish newspapers.

In short, we were faced with two sets of problems: to choose which basic approach to adopt for the course, and to recognise that if approach (3) was chosen then none of the available material seemed appropriate for the purpose.

Method

Before deciding which approach to adopt, it was necessary to investigate which online information retrieval support and resources were available for the journalists working at newspapers in Denmark. Therefore, the largest newspapers in Denmark were contacted. It was decided to only include the largest newspapers in the study. The reason for this was: firstly, the assumption that the large newspapers will be the major future employers for the students; secondly, that many of the small newspapers do not themselves produce articles relying on a large amount of research, context or fact checking. The smaller papers are mainly local papers dealing with local news. Finally, the larger newspapers were assumed to be in front in terms of use of information retrieval sources as well as in information retrieval support and resources provided to the journalists. The definition of a newspaper used in this article is that the newspaper should be published at least five days a week. The reason why such a definition is necessary is that several Danish newspapers publish both a Sunday and a daily issue. Most of the large Sunday newspapers are a special version of the daily newspaper. Some of the daily newspapers, however, do not have a Sunday issue. And finally some of the Sunday newspapers are only issued as Sunday newspapers. It also has to be a 'newspaper': that is a paper that promulgates news and not, for example, commercial tenders. In terms of defining what was meant by the 'largest', it was decided to include only the newspapers that had a weekly net circulation larger than the mean value.

According to DRB [7, p. 181–195] there are 42 newspapers in Denmark that meet the definition. Several of the newspapers have a higher net circulation number for their Sunday issue than for their daily issue. Therefore, in order to decide which are the largest ones, it was necessary to calculate the total amount of circulation per newspaper per week. The total amount of newspapers circulated in Denmark per week is 11.547.436 (1998). The largest newspaper has a net circulation of 1.595.312 per week, while the smallest one circulates 10.224 per week only. The mean value of circulation per week is 274.939. Only nine of the 42 newspapers have a circulation number greater than the mean. These nine newspapers were selected for the study. Table 1 presents the nine newspapers and their net circulation.

Having selected these nine newspapers, the next step was to investigate the availability to the journalists of online information retrieval support and resources. A newspaper was considered to provide online information retrieval support, if at least one of the employees assists with online searching. The employee might be a librarian, a system manager, or a fellow journalist. However, assisting with information retrieval had to be included in the formal job description. Concerning the online information resources, it was decided to define an online information resource as an online service for information retrieval if it had been subscribed for. Thus, for example, 'Dialog' or the 'Internet' was considered *one* online information resource.

Firstly, the newspapers were asked if they had a library, and whether online information retrieval resources were available there (national and international). Secondly, they were asked if a librarian or other professional among the staff was in charge of carrying out online information retrieval. The next questions related to whether or not the journalists themselves had online access and to which of the resources and finally, whether the journalists themselves in fact did carry out searches either in the library or from their desktop.

Findings

All of the newspapers, except one (*Arhus Stiftstidende*), claimed to have a library in accordance with the definition of this article, and all of them, except one (*JydskeVestkysten /Kolding Folkeblad*), had at least one librarian among the staff. For one of the newspapers the librarian was not carrying out searches for the journalists (*Fyens Stiftstidende*). In terms of available online information retrieval resources Table 2 presents how many different resources could be accessed from the library.

Table 2 confirms the assumption that the larger the newspaper, the more information retrieval resources are subscribed to. There seems to be one exception only from that pattern, as *Politiken* subscribes to the same amount of information retrieval resources as does *Ekstra Bladet*. The reason why *Ekstra Bladet* and *Politiken* subscribe to so many information retrieval resources is due to the fact that these two newspapers share their online information retrieval resources, and that they furthermore serve as agents for several of the information retrieval resources in Denmark.

Table 3 presents the results of the study on which online information retrieval

Table 1. The nine largest Danish daily and Sunday newspapers, net circulation and total weekly circulation (1998)

Name of newspaper	**Net circulation**		**Total weekly**
	Daily	**Sunday**	
(a) Ekstra Bladet	222,697	259,130	1,595,312
(b) Morgenavisen Jyllands-Posten	175,687	270,349	1,324,471
(c) Berlingske Tidende	154,687	201,475	1,129,597
(d) Politiken	147,153	197,205	1,080,123
(e) B.T.	138,263	138,263	967,841
(f) JydskeVestkysten/Kolding Folkenblad	95,663	111,103	685,081
(g) Aalborg Stiftstidenende	70,333	85,522	507,520
(h) Fyens Stiftstidende	66,147	94,276	491,158
(i) Arhus Stiftstidende	62,780	79,183	455,863

Table 2. Online information retrieval resources available at the library of the newspapers

Name of newspaper	**Subscription to online information resources**		**Total**
	National	**International**	
(a) Ekstra Bladet	9	16	25
(b) Morgenavisen Jyllands-Posten	10	4	14
(c) Berlingske Tidende	4	1	5
(d) Politiken	9	16	25
(e) B.T.	4	1	5
(f) JydskeVestkysten/Kolding Folkenblad	3	1	4
(g) Aalborg Stiftstidenende	3	4	7
(h) Fyens Stiftstidende	3	4	7
(i) Arhus Stiftstidende	2	1	3

Table 3. Online information retrieval resources available at the desktop of the journalists

Name of newspaper	**Subscription to online information resources**		**Total**
	National	**International***	
(a) Ekstra Bladet	8	1	9
(b) Morgenavisen Jyllands-Posten	5	1	6
(c) Berlingske Tidende	2	1	3
(d) Politiken	8	1	9
(e) B.T.	2	1	3
(f) JydskeVestkysten/Kolding Folkenblad	3	1	4
(g) Aalborg Stiftstidenende	4	1	5
(h) Fyens Stiftstidende	4	1	5
(i) Arhus Stiftstidende	2	1	3

***The Internet**

resources are available to the journalists themselves.

As can be seen the newspapers are restrictive in terms of allowing journalists themselves to search the large international databases except for the Internet. Several of the newspapers had a company policy stating that the journalists themselves could only have access to the free or national services. Such policy was made for economic reasons. The journalists are not in any way trained in information retrieval and, therefore, giving them access to, for example, Dialog or FT Profile was assumed not to be cost efficient. The librarians were considered to be able to perform the searches much more efficiently. Furthermore, three of the newspapers said that some of their departments have their own independent budget and that they subscribe to other services than the ones men-

Table 4. Frequencies of subscriptions for various information retrieval resources by the newspapers: national and international resources

National information retrieval of resources	Newspapers subscribing	International information retrieval resources	Newspapers subscribing
Polinfo (H)	8	Internet	9
Berlingske Avisdata (T)	8	FT Profile	8
Borsen (T)	5	Dialog (H)	3
Aktuelt	5	Reuter Business Briefing (F)	3
Public-com (F)	5	Affairs Data (T)	3
Greens Directory (F)	3		
Artikelbasen (R)	3		
Danbib (R)	3		

H: Database host, T: Full-text database, F: Fact-database and R: Reference database

tioned here. These services are mainly trade directories and encyclopaedias. These are not included in Table 3. It is, however, clear that the journalists are expected to carry out some searches themselves (the cheap ones in the local databases) and to ask for assistance from the librarian for other searches (in expensive international databases).

Table 4 presents the frequencies of subscription for the various databases. The table includes information retrieval resources subscribed to by more than two newspapers only. The reason for this was again, that two of the newspapers share resources.

As can be seen, the newspapers subscribe to all three types of information retrieval resources: full-text, fact-databases and reference databases. With regard to the nternational resources, the Internet and FT Profile are very important. Among the national ones Polinfo and Berlingske Avisdata are he most significant. Polinfo can best be explained as the 'Danish FT Profile': It is a database host that includes several Danish newspapers in full-text but also some fact-databases, while Berlingske Avisdata is the ext archives of *Berlingske Tidende*. Berlingske Avisdata, Børsen and Aktuelt are full-text newspaper databases. Publi-com is a fact-database that includes company budgets and account balances from all major Danish companies. Greens Directory is a local company directory. Artikelbasen is a reference database, which contains references to articles from 750 Danish newspapers and journals and finally Danbib is a reference database that includes references to all documents in Danish libraries.

In summary, Polinfo and Berlingske Avisdata are the national information retrieval resources that are subscribed to by most newspapers, while FT Profile and the Internet are the most important international resources. It is important to note, however, that the newspapers also subscribe to national fact-databases and national reference databases.

Discussion

The study has shown that most of the newspapers have a library and that they have a librarian, who is expected to carry out or assist with searches for the journalists. This finding supports the view that the students should be taught to express their information needs rather than perform searches themselves. On the other hand, the journalists are also to a certain extent expected to carry out searches themselves: they have access to some, mainly local, online information retrieval resources. This finding supports the view that students need practical hands-on training in using some information retrieval resources.

The findings of this study, therefore, suggest that the students should be taught the basic theoretical problems involved in information retrieval, but also that they should be trained in using certain information retrieval resources. This means that approach (3) should be adopted as the basic concept for teaching information retrieval to students of journalism.

Since the students of journalism can expect to have online information retrieval support, the most important topic to teach them was considered to be how to express an information need. Such skills will be useful to them when searching with or without support. The first aim of the course is, therefore, to teach the students how to express information needs as a question. This helps the student to choose between relevant and non-relevant information retrieved: the relevant information is the information that provides part of or the complete answer to the question. The question, however, is not in itself a search statement. The next step, therefore, was to introduce the basics of Boolean logic, Common Command Language, truncation, organising knowledge (controlled and uncontrolled vocabularies), the essentials of how indexing in databases works, and basics about the differences between reference, full-text and fact-databases and database hosts. These topics were selected because they were considered to be general to all information retrieval resources today and in the future.

Following the theoretical introduction, the theory has to be applied to the information retrieval resources that were found to be the most widely used by the newspapers in Denmark. The study has shown, that both the Internet, full-text, fact-, and reference databases are used. And also that both national and international resources are available to the journalists. Therefore firstly, Polinfo, the most frequently used national online information retrieval resource, was selected; secondly, FT Profile was selected for the course. These two databases were considered representative for teaching the students to work with full-text retrieval on a national and international level. Secondly, the students will work with fact-databases. Here company directories, statistical databases and encyclopaedias will be presented. Thirdly, concerning reference databases the students will learn how to use the local Danish 'Artikelbasen' which contains references to journal and newspaper articles. Finally, the students will be introduced to some of the different types of search engines found on the Internet. Manually indexed: for example Yahoo! And automatically indexed: for example Lycos.

Regarding the introduction of different online information retrieval resources, the students will be given exercises that simulate real life situations. On the basis of a news telegram they will be asked to formulate a question, to develop the appropriate search statements on the basis of the question, and finally, to retrieve the information that is relevant in order to answer the question. This requires a targeted textbook written for journalists. This textbook should consist of two parts. Firstly, a theoretical introduction which should be kept at an absolute minimum. Secondly, it should include a user manual for the information retrieval resources selected. For this purpose a new concept had to be invented: the critical user manual. Most user manuals for information retrieval resources do not inform the user of how to manoeuvre between controlled and uncontrolled vocabularies of the retrieval tool in order to optimise a search in terms of recall and precision Neither do they inform about lack of indexer consistency and errors nor how to detec inconsistencies and errors in the indexing.

Conclusion

In terms of selecting the basic approach to teaching information retrieval to students of journalism it was decided to teach the students the basic theoretical problems involved in information retrieval and to apply this theory to exercises on various information retrieval resources. It was decided to include a theoretical introduction to the course for a least three reasons. Firstly, the study o Nicholas and Martin [1] has shown that journalists need to learn how to express thei information needs. Secondly, there is a grea difference in the availability of information retrieval resources at Danish newspapers, which means that it would not be possible to teach them how to use all of them. Finally, that such a basic theoretical introduction, it is hoped, can help the students to cope with future systems.

When students have learnt the basics of information retrieval, they will then be taught how to work with a selection of full-text, fact- and reference databases. These will be selected from the information retrieval resources most frequently used by the largest Danish daily and Sunday newspapers. They will also be introduced to the Internet. For that introduction, however, the focus will be placed on the information retrieval resources available on the Net. In this way, we hope to have provided the students with a general knowledge about information retrieval and the practical skills that enable them to find the information they need – also in the year 2001 when they graduate. By then this investigation will be followed by a comparative study comparing firstly, the information retrieval skills of the students of journalism that have graduated from this University to the students of journalism having graduated from the two other educational institutions (Danmarks Journalisthøjskole and Roskilde University) and comparing secondly, the information retrieval skills of the students of journalism to students of other disciplines here at the University of Southern Denmark.

References

1. Nicholas, D. and Martin, H. Assessing information needs: a case study of journalists. *Aslib Proceedings*, 49(2), 1997, 43–52.

2. Belkin, N.J. Ineffable concepts in information retrieval. In: Sparck Jones, K. ed., *Information retrieval experiments*. London: Butterworth, 1981, 44–58.

3. Harter, S.P. Psychological relevance and information science. *Journal of the American Society for Information Science*, 43(9), 1992, 602–615.

4. Hersh, W. Relevance and retrieval evaluation: perspectives from medicine. *Journal of the American Society for Information Science*, 45(3), 1994, 201–206.

5. Nicholas, D. and Williams, P. *Journalism and the Internet – the changing information environment – pre-fieldwork literature review*. 1998. http://www.soi.city.ac.uk/~pw/ji_lit.htm (visited 12 January 1999).

6. Harter, S.P. *Online information retrieval – concepts, principles and techniques*. Bloomington: Academic Press, 1986.

7. DRB. *Media Scandinavia 1998*. Copenhagen: Danish Association of Advertising Agencies, 1998.

Notes

1. To avoid clumsy constructions resulting from trying to refer to two sexes simultaneously, I have used male pronouns exclusively. However, my information seekers are as likely to be female as male and I mean no disrespect to *men* by my usage. See also Harter [3, 603].

CONTENTS

51–57 A distant mirror?; the Internet and the printing press
David Bawden and Lyn Robinson

58–75 A proposed information environment for enhanced integrated access and value-adding to electronic documents
Chern Li Liew, Schubert Foo and K. R. Chennupati

76–82 In the catalogue ye go for men: evaluation criteria for information retrieval systems
Julian Warner

83–90 Teaching infopreneurship: students' perspectives
Adeline du Toit

A distant mirror?; the Internet and the printing press

David Bawden
Department of Information Science, City University, Northampton Square, London EC1V OHB
dbawden@soi.city.ac.uk

Lyn Robinson
School of Library, Archive and Information Studies, University College, London

The introduction of the printing press to Europe in the mid-fifteenth century, and its effect on the communication of information, are considered, largely by reference to the writings of Elisabeth Eisenstein. Analogies and similarities with the impact of the Internet are identified, as a way of gaining insight into the current communications revolution.

Introduction

The advent, and wide uptake, of the Internet since 1990 has ushered in what is generally regarded as a revolution in the communication of information and knowledge. While this revolution is in progress, its true significance is difficult to estimate. One way of gaining some insight into the matter may be to consider comparisons between the Internet and previous innovations, which have had drastic consequences for the transmission of knowledge.

The earliest two such innovations – language and writing – lie at such a distance in time from us that it is difficult to make any useful comparisons or draw lessons. Of the former we know little or nothing, though many intriguing speculations may be made regarding the evolution of language (see, for example, Pinker [1] and Deacon [2]). Of the origins of writing, more is known [3, 4], and comparisons may be made with those current cultures that have no written script. But in neither case can we examine the change wrought on the information culture of the time by the innovation, still less hear the views of the participants.

Conversely, a great deal is known about the most recent innovation, or rather complex of innovations, which characterised the communication and information revolutions of the nineteenth century. These innovations included mass publication of books and periodicals, steam transportation, national postal systems, electronic communication by telegraph and telephone, public and special libraries, bibliographic classification schemes and cataloguing rules, abstracting and indexing services, and many of the current systems of business document and record handling. They are summarised by Benninger [5] as a 'Control Revolution', whose effects still largely provide the framework – intellectual and, to a surprisingly large extent, technical – within which current information systems and services operate. But, precisely because of the multi-faceted nature of this 'revolution', it is hard to know which elements may be selected for comparison with the Internet; to select any one is certainly to ignore other highly significant, and certainly interacting, factors.

This leaves one remaining innovation, whose effects are certainly comparable to those of the Internet: the introduction of printing within Western Europe in the latter part of the fourteenth century. This innovation, and its results, has been comprehensively reviewed by Eisenstein in a two-volume work [6], and an abridged edition [7]. This authoritative work is taken as the basis for a consideration of what may be learnt by a comparison of the printing press with the Internet. The comparison is possible because we are dealing with the ramifications of a clearly defined technical innovation; and because the careful scholarship of Eisenstein and many others has thrown up a remarkable amount of information on the effects of the printing press.

The comparisons made here are selective rather than comprehensive, are impressionistic rather than rigorous, and utilise rough analogies rather than detailed metrics. The Internet is not the same as the printing press, nor will its effects be the same; the interactivity and dynamic nature of the Internet, and its intermingling of formal and informal communication, ensure this. But it may be that there is sufficient commonality to create some interesting, and perhaps insight-provoking, ideas. Drawing too strong an analogy between present and past times is a danger in all historical studies, but the success of works such as that on which the title of this paper is based [8] suggest that it is a risk worth taking.

The print revolution

Eisenstein's main theme is the change brought about by the advent of printing on the information culture (though she does not use the term) of the time; from the age of scribes to the age of print. (For background on the scribal culture, see Reynolds and Wilson [9], or for a shorter account de Hamel [10].) She looks beyond the evident fact that a larger number of works became available in much greater quantity than beforehand, to consider the impact of the printing press on the structure of knowledge, and of thought itself.

That it was a true revolution, rather than a slow evolution, is worth emphasising. After the first establishment of printing workshops in the Rhineland in the 1450s, the craft spread throughout Europe rapidly. By the end of the century, about eight million books had been produced. While comparisons with the production of hand-written texts by scribes are extremely difficult, due to the limited information for earlier centuries, and the difficulty of defining what constitutes a book prior to the print era, this figure must correspond to several centuries at least of scribal production. One estimate is that at the time of Gutenberg, about 30,000 manuscripts contained the world's store of recorded information, to be compared with 1.25 million titles printed by the time of Shakespeare, 150 years later; truly a 'bibliographical explosion' [11, p. 75].

It is worth noting immediately that we tend to think that print equals text. However, as Eisenstein reminds us, we would do better to think of printing and engraving: the two technologies developed together. She quotes the view of the American scholar George Sarton that it is best to consider 'a double invention: typography for the text, engraving for the images'.

Engraving provided not merely illustrations, but mathematical formulae, scientific diagrams, and the like; thus by the end of the fifteenth century, letters, numbers and pictures were all amenable to reproduction, with the printed book allowing interplay between all the elements. While it would be trite to regard this as a kind of Renaissance multimedia, it is as well to remember that technologies for integrating text and graphics have been available for a considerable time. The simplistic idea that printing moved the communication of information from graphic image to text, and that the Internet and similar media are redressing this balance, is not borne out by several strands of evidence presented by Eisenstein; not least the fact that the immediate impact of printing/engraving was to make recorded images much more readily available.

Of course, one distinction between the printing press and the Internet may seem immediately evident. Printed books were aimed, initially at least, at a small and elite market, whereas the Internet is accessible to a mass audience. Though literacy rates through history are notoriously difficult to assess, Eisenstein presents evidence to show that they may have been higher than generally assumed; but clearly the proportion of people who could afford to purchase significant numbers of printed books would not have been large. However, the distinction may not, in fact, be so marked. It is clear that the advent of the printed book affected the lives, thoughts and beliefs of many more than those who acquired private libraries.

Eisenstein refers to a 'knowledge explosion', accompanying the 'Northern Renaissance' of the fifteenth century, with the literature of the period, involving writers from Marlowe to Rabelais speaking of the 'intoxication' of the availability of written knowledge. In fifteenth century England, the printers and booksellers catered 'to the needs of lowly bakers and merchants as well as to those of lawyers, aldermen or knights [7, p. 32]'.

It seems that printed books were used by a rather wider cross-section of the population than might have been thought, and that their availability had an influence well beyond the scholarly elite. This may not be so far removed from the situation of the Internet, in that many people who have never used the Internet know of it, have opinions about it, and are aware of its power to change the information landscape with which they are familiar.

Paradoxes of print

In two respects, Eisenstein shows how the printing press, far from creating an entirely new form of information product, built on, and extended, what was already present, with sometimes paradoxical results.

For one thing, many of the concepts and techniques of information design, which were known to the scribes of the time, found much wider expression in print. Most strikingly, alphabetical order only became widely understood from the sixteenth century, though the alphabet itself had been devised many centuries before. This change may be attributed to book indexes and catalogues of books, themselves the results of the adoption of printing [7, p. 64-73]. When each copy of a work had to be produced by hand, the index, title page and similar aids would be produced to meet the individual, and idiosyncratic, needs of each reader; only with mass production of uniform texts did the need for standards in such matters become important. Similarly, when libraries and booksellers had a small stock of individual manuscripts, the products of scribal industry, there was again no need for the standardised form of arrangement of a catalogue. But:

> after the advent of printing, however, shelf lists were supplemented by sales catalogues aimed at readers outside library walls, while any index compiled for one text could be duplicated hundreds of times [7, p. 66].

Printing promoted not only standard forms of cataloguing and indexing, but also systematic use of such aids to the reader as footnotes, cross-references, page numbering in Arabic numerals, consistent title pages in works, tables of contents, links between text and figures, and so on. All of these elements had been known in scribal culture; indeed, features such as tables of contents, cross-referencing and page numbering date from the introduction of the codex, the earliest form of a book with individual pages rather than a continuous scroll, in late Roman times. They could find full application, however, only in a world of print. Of course, there is a linkage between the two, in that catalogues of whatever kind are immeasurably easier to construct, given consistent tables of contents; an idea which has a parallel in modern metadata issues.

This element of using the new technology to enable the instantiation of ideas devised at an earlier time finds some parallels with the Internet, most obviously in its use of the ideas of hypertext, devised by scholars of earlier times, such as Vannevar Bush [12] and Ted Nelson [13]. Rada [14] and Gilster [15] argue that the origins of hypertext are to be found in those early manuscripts, which allowed footnoting, cross-referencing, and page indexing, to give connections between different parts of the text. In this view, printing simply extended the applicability of the idea, allowing links to other texts. Because of the standardisation imposed by the printing process, a reader could follow a link to any copy of the cited text, with confidence that they would all provide identical information; this would not have been possible in the era of hand-written manuscripts. Its digital implementation simply makes the process faster, and allows immediate access to linked documents.

The second point in considering the relation of printing to the products of the older scribal culture concerns the nature of what was printed. For one thing, while the usual assumption is that the printing press was used to promote humanist texts, and the earliest works of modern science, it is clear that, for the first few decades of printing at least, older works outnumbered new writings. Classical and Arabic texts were accompanied by arcana, occultism and 'fraudulent esoteric writings', so that:

> we must not think only about new forms of enlightenment when considering the effects of printing on scholarship. New forms of mystification were encouraged as well [7, p. 48].

The initial effect of printing, even in the academic sphere, was to greatly expand the holdings of libraries, but with material of very diverse quality: an assessment and sorting came later. Something of the same pattern, albeit on a very much reduced time-scale may be seen in the emergence of subject gateways to bring quality assessment, and access by standard means of resource description, classification and indexing, to the burgeoning mass of Internet material [16].

Nor were more popular subjects neglected:

> There is simply no equivalent in scribal culture for the avalanche of 'how to' books which poured off the new presses, explaining by 'easy steps' just how to master diverse skills, ranging from playing a musical instrument to keeping accounts [7, p. 64].

Even current concerns over the effect of the Internet in making pornography, and other salacious material more readily available was mirrored in the early decades of print, with printers enthusiastically providing 'scandal sheets, "lewd Ballads", "merry bookes of Italie" and other "corrupted tales in Inke and paper" [7, p. 94]'.

Print culture, over the centuries, has evolved sophisticated forms of scholarship, divided and sub-divided into disciplines and schools of thought, each with its own methods and standards. But this was by no means so in its early decades, or even centuries. The extent to which the interests of the early modern scientists overlapped into subjects such as alchemy, astrology, and mystical numerology is striking, Sir Isaac Newton being only the most obvious example. In its early years, print culture as a whole was characterised by a 'somewhat wide-angled, unfocused scholarship [7, p.45]'. This should be borne in mind, when considering some of the complaints of traditional scholars against an over-reliance on networked resources; perhaps every new communication medium must pass through such a stage.

The wandering scholar and the digital nomad

One aspect of the early years of printing strikes a close note of similarity with the Internet: the way in which distance is overcome in giving access to information. The centuries between the establishment of the first universities and the widespread adoption of printing were the time of the 'wandering scholar', travelling between universities, monasteries and other places where manuscripts were to be found. Travelling was absolutely necessary, when the scribal system meant that it was unlikely that copies of a range of material in any discipline could be found in any one place. Printing meant that universities and other institutions, even private individuals, could possess the major corpus of a discipline in one place; the era of the 'sedentary scholar' was established. The analogy with the ability of the Internet to bring the corpus to a still more focused location – the user's desk – seems clear. Of course, the Internet, with other digital technologies, facilitates communication and information access for mobile users. (See, for example, Makimoto and Manners [17].) But, rather than any kind of reversion to the wandering scholar, this can be seen as simply an extension of the printing press's ability to provide knowledge to its users wherever they may be; information comes to the user, rather than the user having to travel to the information, but in a fuller manner.

McGarry [18] points out that the 'invisible colleges' of scientists, stemming from the seventeenth century, were made possible only by printing, making possible the scientific journal as a form of communication, and by the development of the first person-to-person postal services. He suggests that subsequent advances in communication (to the time of writing in the mid-1970s), particularly the telephone, air mail, and long-distance passenger air travel, enabled members of such colleges to communicate, and meet in person, much more rapidly than was possible by the conventional publication process. Subsequent developments in electronic communication via the Internet have speeded up the communication process still further, arguably minimising the need for physical presence and transfer of paper documents. This may be seen as another example of the Internet bringing out the same qualitative change as the printing press: enabling person-to-person interaction among specialised interest groups, and allowing for a subtle mix of formal and informal interaction.

Quality and repeatability

Another effect of printing was to greatly diminish the role of 'commentator and glossator': those scholars who, for want of access to a range of texts, devoted their whole lives to analysing and re-analysing a single text. Though it is unlikely that such could be found today, it is not unreasonable to suggest that the highly specialised forms of scholarship which thrive in a print-based academic culture may be under threat from the ability of the Internet to deliver a wide variety of relevant material outside the immediate domain as expressed by, for example, a printed scholarly journal.

This ties in with the point made above – that the early years of print were characterised by the printing and distribution of many copies of elderly texts. Yet it was the very availability of the texts for comparison and cross-checking, Eisenstein argues, that stimulated the emergence of new forms of knowledge. This is shown by the examples of planetary theory within astronomy, botanical knowledge as expressed in herbals, and cartography. For the latter, the atlases of Abraham Ortelius, which went through many editions allowing the older erroneous maps to be compared and supplanted. The very concept of having a standard and repeatable version of a text (or map or diagram), which can be corrected and improved in successive editions, of course, is a feature of printing/engraving which was impossible in a scribal system of hand copying of manuscripts. This led to a variety of tools for data compilation and comparison, including dictionaries, lexicons, concordances and the like, which form the basis for modern print scholarship, but which existed only in rudimentary form before Gutenberg. The development of equivalent tools for Internet information will be a major step in marking the maturity of the new medium.

We should also note one rather paradoxical feature of this process, identified by Eisenstein: an initial degradation in the quality of books, ranging from scripture texts to botanical herbals. In the hands of printers wishing simply to produce rapid profits, the errors, which had always been present in manuscript copies, were replicated, and in some cases increased. (Nor could an individual scribe commit the kind of 'standardised error' which created the 'wicked Bible' of 1631, when the word 'not' was omitted from the Seventh Commandment.) However, the much greater availability of printed copies from varied sources meant that such errors could rapidly be identified and corrected, so that the ultimate result was a higher overall quality. There seems to be an echo here of concerns about the quality of Internet materials produced without the quality control processes of conventional print-based publishing, raised in the intervening centuries to a high level of sophistication.

Not only was the content of many printed works to some degree archaic, but so was the form and style of their production. The choice of script was carefully chosen, in many cases, to imitate handwriting, decorations equivalent to those added by scribes were included, while erasers and paintbrushes were applied to the final product to give an imitation of a hand production [18, 14]. It may therefore not be surprising to find how many Internet resources are simply screen-based versions of what might be expected on the printed page. The development of information with novel structure and design to suit the capabilities of the new medium is slow in coming; as with printing, this is due to a combination of lack of imagination on the part of information providers, with a perception that users will be happier with the kind of product with which they are familiar.

Other features of the introduction of printing strike surprisingly modern chords. There is a concern for preservation of material over time. On the one hand, there was general belief that paper would not last as well as parchment, so that printed material would be necessarily more ephemeral. On the other, there was a rapidly increasing recognition that the making and dispersal of multiple identical copies was the best defence against the loss of material contained in a few manuscripts, however carefully guarded. Here we may see a precursor of current concerns about digital preservation, and also the feeling that something placed on the Internet is, in a way, invulnerable to censorship. Of course, the origins of the Internet were in the US defence agencies' ARPA network, which relied on the concept of decentralised passage of messages through a redundant network; not without an analogy in the preserving role of multiple print copies.

Social aspects of the new technology were not ignored. 'Complaints about the 'sullen silence' of newspaper readers in seventeenth-century coffee houses point to the intrusive effect of printed materials on some forms of sociability' [7, p.94] appear to echo current concerns about the isolating effects of over-reliance on the Internet. Similarly, Eisenstein's identification of the amplification and reinforcement of 'myths and clichés', often stemming from ancient scribal sources, but gaining much deeper roots in print, seems to have something in common with the modern phenomenon of Internet myth and spoofs (see Gilster [15, Ch. 4], and Crawford [19, Ch. 11]).

Conclusions

It seems clear that many of the issues and consequences surrounding the introduction of the printing press, as it succeeded scribal manuscript production, have a similarity to those noted as the importance of the Internet increased.

We should not expect too exact a correspondence, for reasons noted above. Also, while print rapidly replaced manuscript, the same is not necessarily true for Internet and print. Gilster, for example, argues that networked information sources extend, rather than supplant, the power of print; 'the two technologies intertwine like DNA strands, the double helix of the twenty-first century's intellectual revival' [15, p. 26].

Nonetheless, examination of an earlier communications revolution may provide lessons for understanding the one currently under way. McGarry [18] argues that the fundamental relationship between a medium of communication and the one that it replaces is an initial dependency of the new on the old form, and some of the issues discussed above support this.

Crawford [19] suggests an alternative general rule: that new media complement rather than replace the old. However, with respect to the printing press, his comparison is with oratory, rather than with the written scribal manuscript. Oratory has not been supplanted entirely, though the nature of spoken communication has been changed dramatically by the availability of printed materials, whereas the hand-copied text has not been a viable information product for some hundreds of years. Our view of the impact of the Internet on the world of print may, if we follow Crawford's thought, depend on whether we believe that it is a new way of presenting the written word, in which case print may be expected to wither as did the scribal manuscript, or whether we see it as an entirely new medium of communication, in which case it may be expected to find a niche alongside speech and print.

Gilster offers a rather broader perspective:

> I see the Net as having the same effect all technology does. It offers new possibilities that have to be considered within the context of an unchanging human nature [15, p. 22].

Examining earlier technological innovations, where their significance can be appreciated over the gap of time, may be a way of assessing this. The printing press may indeed offer a reflection of the nature and significance of the Internet.

References

1. Pinker, S. *The language instinct.* Harmondsworth, Middx.: Penguin, 1995.

2. Deacon, T. *The symbolic species.* Harmondsworth, Middx.: Penguin, 1998.

3. Hooker, J.T., ed. *Reading the past.* London: University of California Press/British Museum, 1990.

4. Robinson, A. *The story of writing.* London: Thames and Hudson, 1995.

5. Beninger, J.R. *The control revolution; technological and economic origins of the information society.* Boston: Harvard University Press, 1986.

6. Eisenstein, E.L. *The printing press as an agent of change.* Cambridge: Cambridge University Press, 1979.

7. Eisenstein, E.L. *The printing revolution in early modern Europe.* Cambridge: Cambridge University Press, 1983.

8. Tuchman, B.W. *A distant mirror; the calamitous fourteenth century.* Harmondsworth, Middx.: Penguin, 1980.

9. Reynolds, L.D. and Wilson, N.G. *Scribes and scholars.* Oxford: Oxford University Press, 1968.

10. Hamel, de C. *Scribes and Illustrators.* London: British Museum Press, 1992.

11. McGarry, K. *The changing context of information.* 2nd ed. London: Library Association, 1993.

12. Bush, V. As we may think. *Atlantic Monthly,* 176(1), 1945, 101-108.

13. Nelson, T.H. Opening hypertext: a memoir. In: Tuman, M.C., ed. *Literacy Online: the promise and peril of reading and writing with computers.* Pittsburgh: University of Pittsburgh Press, 1987, 43-57.

14. Rada, R. *Hypertext: from text to expertext.* Maidenhead: McGraw Hill, 1991.

15. Gilster, P. *Digital literacy.* New York: Wiley, 1997.

16. Robinson, L. and Bawden, D. Internet subject gateways. *International Journal of Information Management,* 19(6), 1999, 511-522.

17. Makimoto, T. and Manners, D. *Digital Nomad.* Chichester: Wiley, 1997.

18. McGarry, K. *Communication, knowledge and the librarian.* London: Clive Bingley, 1975.

19. Crawford, W. *Being analog; creating tomorrow's libraries.* Chicago: American Library Association, 1999.

A proposed information environment for enhanced integrated access and value-adding to electronic documents

Chern Li Liew, Schubert Foo and K.R. Chennupati
School of Applied Science, Blk N4 #2A-32, Nanyang Avenue, Singapore 639798
PH8598467@ntu.edu.sg

Providing enhanced access and added value to electronic documents (e-documents) will require interfaces that effectively mediate between the information seeking needs of the users and the information that the e-document has to offer. A proposed information environment (PIE) to support effective and creative use of e-documents to fulfil users' various information tasks and needs is presented. Providing a suite of novel features and interactive tools that can be flexibly combined, the PIE allows users to apply multiple novel ways to intuitively query and navigate information in an e-document. The querying and browsing processes in the PIE are supported by various interactive and visualisation techniques. Users work within a visually sovereign, integrated environment for information gathering and organising, based on navigable, fractional information objects that are also affiliated with rich metadata and additional layers of value-adding information. This paper describes the conceptual design of the PIE and uses a representative scenario to show how the PIE is likely to be used for interacting with electronic journals (e-journals) and the ability of the proposed environment to provide enhanced user interaction and value-adding.

Introduction

The exciting new possibilities posed by the growth of electronic media collections and digital libraries bring with them a new set of challenges to overcome. While there are undoubtedly technical problems associated with storage and delivery of digital media that have yet been solved, some of the most challenging problems are less obvious. Electronic collections of documents are fundamentally changing the way users interact with a document. When a document is prepared, stored and presented electronically, the potentials for integrating, categorising and offering the data it contains are greatly enriched. For instance, users need new means to locate specific information to get an overview of the structure and layout of a document, and to browse through the different sections within the document. Electronic documents (e-documents) have also made possible manipulation of information objects in novel multiple complimentary ways. Users no longer approach a document as passive recipients. Instead they need to be active composers and manipulators of information to fulfil their various information needs and to get additional values out of e-documents.

New user interface challenges are emerging because users need to explore and perform many diverse tasks involving huge quantities of abstract information contained within an e-document. Effective environments to provide user interface mechanisms that support multiple information gathering and analysis tasks are needed. For instance, interacting with electronic journals (e-journals) might include tasks such as:

- searching for, analysing and narrowing the scope of information to relevant subsets;
- re-organising information to create new knowledge;
- filtering information to focus on relevant and important information;
- controlling the level of details (on-demand) of information in display;
- checking details-on-demand while maintaining the context of overall tasks;

- detecting important patterns and relationships;
- communicating, collaborating, and learning from other users;
- acting directly on information through interactive tools.

Visualising information is one approach to addressing these challenges. However, visualisation must engage much more than merely permitting users to 'see' information. Users must also, for instance, be able to manipulate it to focus on what is relevant, and re-organise it to create new information. Exploration is only the beginning. Users must also share and communicate information in collaborative settings, and act directly to perform their information-seeking tasks based on this information. These goals essentially suggest the need for innovative information workspaces with new interaction approaches.

Earlier studies of enhancing interaction and value-adding to e-documents have mostly focused on isolated solutions. Some of these works have dealt specifically with the information presentation aspect [1, 2]. Others are concerned with providing access to additional information via means such as hyperlinks [3, 4, 5]. Some researches into e-journals [6, 7, 8] are also increasingly addressing the possibilities of transcending the limitation of the printed journal to include unprintable materials – multimedia files that contain audio, video, animation and 3D graphics.

The need to support a wide variety of basic information manipulation and analysis tasks has also given rise to many new interactive techniques, visualisations and applications. Some applications support, for instance, creating new information and visualising relationships among them [9]. Others support controlling the level of aggregation of data and rapid filtering of data subsets [10, 11]. Still others support navigation through large data spaces [12, 13, 14]. A fundamental user interface design question is how one can use the complimentary features of various visualisation and analysis tools in a uniform and co-ordinated manner? Even for just these few example activities, for instance, how can one create new data attributes with one tool, filter the same data with another and visualise the resulting subsets with a third tool?

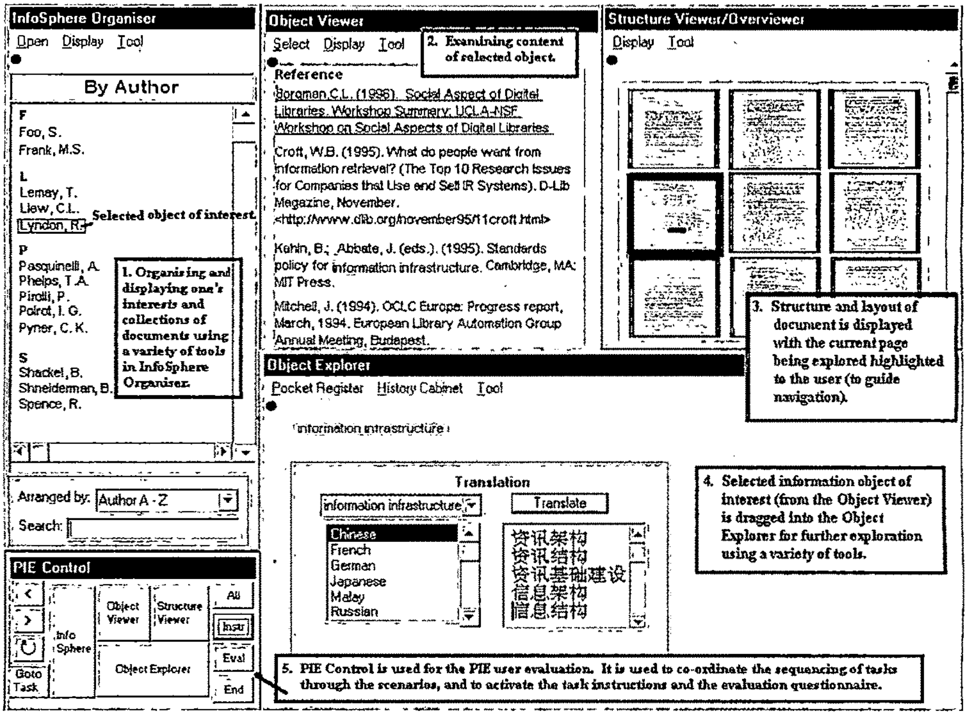
Little research has been carried out on this broader problem of supporting the wide range of information seeking, analysis and communicating tasks within a single user interface environment. This embraces an appreciation of how to create a consistent information workspace where users can make seamless transitions in their utilisation of multiple tools, visualisations and applications. In the earlier phase of our work [15], we explored some design considerations for such an environment, based on our reviews and analyses of user-information interaction needs and requirements, developments in electronic information organisation and management, and of novel user interface technologies. We have identified and presented a set of desired properties for both the interaction environment and the information objects within the environment. These properties form the foundation features for a range of new interfaces that are designed to provide an enhanced level of user-information interaction, and to support e-documents value-adding. This paper describes the conceptual design of a proposed information environment (PIE) that attempts to respond to them.

A central element of our study and approach is to provide users of e-documents with multiple means for expressing their intentions during various information-seeking tasks. 'Multiple means of expression' in this context refers to visualisation and interaction techniques that are useful for different purposes, but are often effective when used in combination as they support representations of different kinds of objects, actions and goals.

The Proposed Information Environment (PIE) interface overview

The PIE is divided into four workspaces. The use of multiple workspaces enables an integrated and unifying environment that allows for visual sovereignty. Figure 1 shows the main interface and the four basic parts of the PIE together with the 'PIE Control'. The PIE Control is not part of the PIE but is used for carrying out the PIE evaluation experiments.

Figure 1. The PIE main interface



Selected object of interest.

1. Organising and displaying one's interests and collections of documents using a variety of tools in InfoSphere Organiser.

2. Examining content of selected object.

Boraman,C.L. (1996). Social Aspect of Digital Libraries. Workshop Summary: UCLA-NSF Workshop on Social Aspects of Digital Libraries

Croft, W.B. (1995). What do people want from information retrieval? (The Top 10 Research Issues for Companies that Use and Sell IR Systems). D-Lib Magazine, November. <http://www.dlib.org/november95/11croft.html>

Kahin, B.; Abbate, J. (eds.). (1995). Standards policy for information infrastructure. Cambridge, MA: MIT Press.

Mitchell, J. (1994). OCLC Europe: Progress report, March, 1994. European Library Automation Group Annual Meeting, Budapest.

3. Structure and layout of document is displayed with the current page being explored highlighted to the user (to guide navigation).

4. Selected information object of interest (from the Object Viewer) is dragged into the Object Explorer for further exploration using a variety of tools.

5. PIE Control is used for the PIE user evaluation. It is used to co-ordinate the sequencing of tasks through the scenarios, and to activate the task instructions and the evaluation questionnaire.

Each of the workspaces serves a specific (yet complimentary) function within the environment. The workspaces operate in an independent yet closely co-ordinated manner. At any given time, users can move information objects form one workspace to any other workspace to perform different tasks to fulfil their various needs. Similarly, users have the autonomy and flexibility to use either all the workspaces in a unified manner, or to use each of them independently according to different needs. They can, for instance, call up only the necessary workspaces, minimise or maximise them, and move the workspaces around in the environment according to their preferences and needs.

We begin the presentation of the PIE with a brief overview of the whole environment and each of its basic components (workspaces). This will be followed by a more detailed description of each of the workspaces and the purposes they each serve. Along the way, we introduce a number of concepts and analogies applied within the PIE. This in turn, will be followed by a scenario (containing a set of mini scenarios) that is extensively intended to demonstrate the use of the various features and tools within the PIE to support a user's various information-seeking needs in interacting with an e-journal.

Basic parts of the PIE

1. **InfoSphere Organiser** – This is the part of a wider information work sphere whereby users organise and display their information collections. These include their personal collections of information, or information available outside the user's current work environment (e.g. collections of journal articles that the user has subscribed to from various digital libraries). Various interaction and information visualisation techniques can be applied here to provide users with a view into this sphere

and the various items in the collections. Users could, for instance, want an overview of the items in a collection to quickly select some relevant items, and to be able to extract the information (objects) that they want to explore further. A search mechanism is in place for the user to quickly locate desired items. Within the *Infosphere Organiser*, users can, for instance, zoom in on the information clusters (e.g. by using a continuous zoom mechanism on a document map view) and navigationally explore the information object collections in detail until they eventually locate the actual information (e.g. an article within a journal collection). To further explore an object, the user simply drags the object and drops it into other workspaces. This broader workspace is closely integrated with the following user's personal workspaces.

2. **Object Viewer** – This is the workspace whereby users get a first glance of the content of the selected object. For instance, once a journal article has been selected and dragged into *Object Viewer* , users will be given an option to either have the first page of the article or only the abstract displayed to get a quick view of the contents of the object. This part of the PIE is closely integrated and co-ordinated with the *Structure Viewer/Overviewer*. The user can make use of the co-ordinated display and directed-navigation to browse through the document. To further explore an information object, the user can select (e.g. the whole article or a diagram within the article) and drop it into the *Object Explorer.*
3. **Structure Viewer/Overviewer** – Next to the *Object Viewer* is where an overview of the object structure is displayed. Users will be given a set of display options to examine the structure and layout of the object. By default, the display would be selected based on the object's inherent structure (i.e. the most suitable display). For example, if the selected object is a set of articles by a particular author, the information may be represented in the form of multiple time line, showing multiple facets of and relationships between the records using selectable icons. If the object under examination is an article, thumbnail pages of the document are displayed. This workspace also offers the users a set of tools to interact with the display. Users can interact with the interface to display the records in multiple representations to explore various patterns and to discover potential new information.
4. **Object Explorer** – This is the 'value-adding' aspect of the PIE that allows users to perform a variety of other tasks to further analyse various information objects. Users could interact with more than one (set of) object simultaneously. The user is able to, for instance, specify one set (e.g. of words) to be used as a query, and another set (e.g. of articles) to be used as the ad hoc collection to be searched on. The system then returns a third set (i.e. matches) which the user can save as a personal collection (in *the InfoSphere Organiser*), or to perform further exploration. Here, users are given a set of widgets (e.g. a set of functional lens, tools to view and add digital comments and annotations, and tools to browse through related works) available from the 'Tool' menu to interact with various information objects. There is also a '*History Cabinet*' – serving as a 'store' for the user to keep track of search and browse interactions, and a '*Pocket Register*' – serving as a 'back pocket' whereby information objects can be deposited temporarily, staying ready-at-hand while the user explores some other objects in the workspace. These two widgets are governed by certain authorisation (i.e. access control) and protection (i.e. security) mechanisms.

The 'body & clone' analogy applied in the PIE

An important operation in the PIE interface is the movement or transfer of information objects throughout the environment. Information objects can be moved directly in a group or individually among the various workspaces. The advantage of having multiple workspaces for interaction in the PIE is that while the user is navigating through an information space with various information objects, the user often needs to focus on certain objects (e.g. in checking out a link provided in a journal article) while still maintaining the context of the whole exploration space (e.g. where the link lies within the original article in its context). During and after the focusing, it is also important for the user to be able to tell where they are in the information space. One method to allow this is to attach a 'handle/switch' (to invoke a set of tools) to an object in its original space. However, some context may be lost in this case as expanding an object may cause other objects in the original space to be occluded. Another method is to make all other objects (other than the selected object) invisible, but this also sacrifices scene context. Yet, another method is to employ a 'body & clone' analogy, where object(s) can be moved to a separate workspace.

In the PIE, the 'body & clone' analogy is applied. In doing so, context is maintained as object 'bodies' are left (highlighted) in the original (home) space while the 'clones' are drawn out and explored in a separate (action) space. Object 'bodies' always appear in its original form and context. A new physic is introduced to objects – an object may pass through other objects in the space and have its position and form altered in a separate space. Using this analogy, users can also move and slide objects (focus) back and forth more intuitively and easily between 'home' and 'action' space. This also allows the user to switch between focus sets in order to quickly locate objects that are relevant to the current task. The user can also interact with more than one object (or set of objects) to, for instance, compare, to reveal pattern, to discover new relationships and doing all these while maintaining context of the whole information space. It also allows users to examine multiple object sets based on non-spatial properties (i.e. they may not be located close to each other in the original space). For instance, users may want to extract a few objects (e.g. links, images) scattered within an article and examine them later in the separate space to tease out new information.

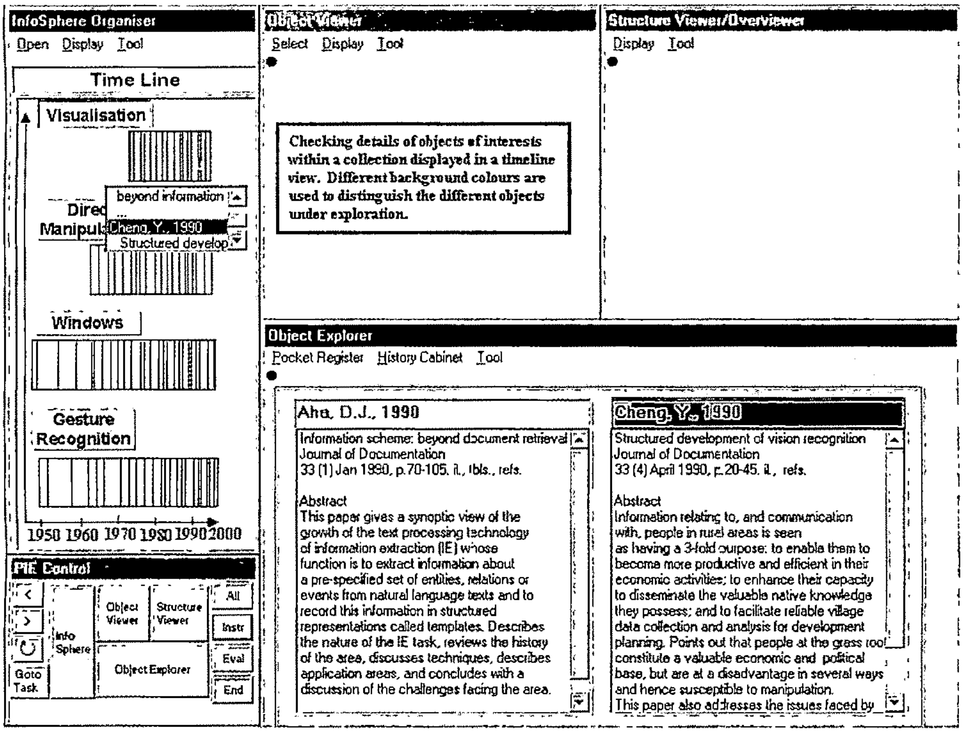
The advantage of the 'body & clone' analogy is obvious. It allows users to interact with any part of the information space at any time, with multiple objects (or sets) and apply various operations to these objects – allowing multiple operations depending on the tools or features selected. The analogy can be extended to include a 'multi-clone' analogy whereby users may have more than one 'clone' of the same object in the 'action' space to perform various queries on the 'clones' to, for instance, compare various sets of query results on the same information object. This 'body & clone' analogy is illustrated in one of the mini scenarios to be presented in the next section of this paper.

At this point, it is appropriate to summarise how the design desideratum [15] are shaping the PIE interface. The single unifying environment with independent yet co-ordinated workspaces and ubiquitous movement of information objects is meant to enable a more integrated environment to support the carrying out of a sequence of related tasks. Direct manipulation is the fundamental mode of interaction. Tools to enhance information gathering and analysis tasks (e.g. navigation via information layout and structure, and infinite zoom) are incorporated into the PIE. The information objects within the environment are fractionally structured and can be laid out on the information surface to depict this fractal structure. All objects (each associated with a rich prescription of metadata) at any granularity are queriable and navigable. Objects collected at any of the workspaces can be temporarily stored into the *'Pocket Register'* that serves as an interactive, multi-object, visible 'back pocket' that helps reduce space contention. The *'History Cabinet'* meanwhile, maintains a record of history for backtracking.

The PIE for interacting with e-journals – a scenario

In order to further examine and convey the use of the PIE for interaction with e-documents, it will be useful to study each of the interface components with regards to the functions that they support. We demonstrate this through a typical scenario and some representative displays in which an Information Science student interacts with e-journals via the PIE to gather relevant and useful information and images for a digital libraries proj-

Figure 2. 'Time Line' view of items within an external collection in InfoSphere Organiser and locating and checking details of objects of interests



Checking details of objects of interests within a collection displayed in a timeline view. Different background colours are used to distinguish the different objects under exploration.

Aha, D.J., 1990

Information scheme: beyond document retrieval
Journal of Documentation
33 (1) Jan 1990, p.70-105. il., tbls., refs.

Abstract
This paper gives a synoptic view of the growth of the text processing technology of information extraction (IE) whose function is to extract information about a pre-specified set of entities, relations or events from natural language texts and to record this information in structured representations called templates. Describes the nature of the IE task, reviews the history of the area, discusses techniques, describes application areas, and concludes with a discussion of the challenges facing the area.

Cheng, Y., 1990

Structured development of vision recognition
Journal of Documentation
33 (4) April 1990, p.20-45. il., refs.

Abstract
Information relating to, and communication with, people in rural areas is seen as having a 3-fold purpose: to enable them to become more productive and efficient in their economic activities; to enhance their capacity to disseminate the valuable native knowledge they possess; and to facilitate reliable village data collection and analysis for development planning. Points out that people at the grass root constitute a valuable economic and political base, but are at a disadvantage in several ways and hence susceptible to manipulation.
This paper also addresses the issues faced by

ect. The student is interested in locating some journal articles with regard to such issues as user interface, search and retrieval, and delivery mechanisms used in digital libraries.

Browsing through collections

The student wants to first check his external collection. The external collection is the student's subscription to journal collections from various digital libraries based on his research needs and interests. He does this in the *InfoSphere Organiser*. The *InfoSphere Organiser* of the PIE provides the functionality to allow him to organise, browse, and highlight relevant aspects of his collection. He can get a quick view of the list of items (e.g. arranged either by subject/topic, author or journal titles) and a search mechanism is available for him to quickly locate and retrieve desired items. There are a number of options to display this collection. For example, the student can seek out works on various user interface technologies and display this information in the form of a time line.

Figure 2 shows how the information can be displayed in a 'Time Line' view – depicting the distribution of work done on various technologies and the period during which this work had been carried out. To check details of work included here, he first pauses his mouse cursor over a point along the time bars. Each time he pauses over a point, a small panel pops up unobtrusively near the point, providing a quick indication of the documents contained within the selected point. He can select an article of interest and check its details. To avoid cluttering the space in *InfoSphere Organiser*, once the student has selected the article, a separate panel pops up in another workspace – *Object Explorer*, displaying more details of the selected article. More than one panel can be displayed, allowing the student to examine and compare different articles simultaneously. Different co-ordinating background colours are used to distinguish different panels and information highlighted (see Figure 2).

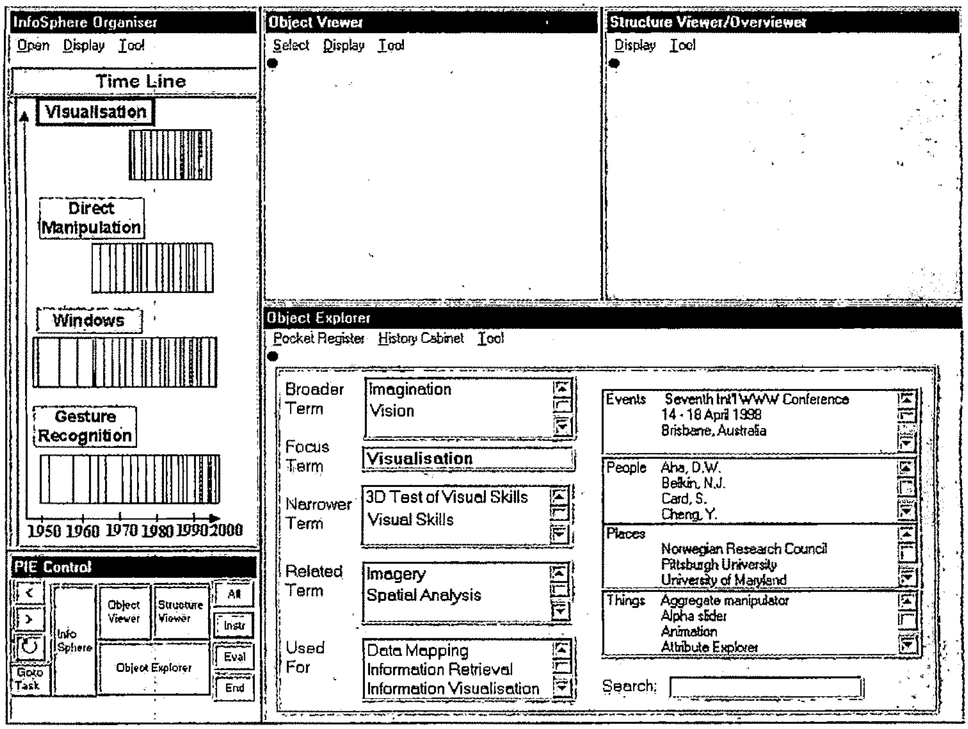
Seamless access to other information sources

The student notices the term 'visualisation' and wishes to find out more about this term. Instead of having to check out the term in, for instance, a printed thesaurus, he can now perform this task directly within the PIE. He highlights the term and drags it into *Object Explorer*. From the 'Tool' menu, he selects the 'Thesauri & Knowledge Package' option. A new panel pops-up, grouping standard relationships and a knowledge package capturing and listing events, people, products, amongst others that are associated with 'visualisation'. The 'Thesauri & Knowledge Package' can be thought of as a tool that groups together terms based on how their corresponding concepts are represented and organised in an intelligent reasoning system. Such a knowledge-rich approach has been found to be very useful to enhance search and retrieval in digital libraries [16]. Figure 3 shows the invoked feature. The student can also call up the 'Glossary/Definition Lens' and the 'Translation' options in *Object Explorer* to get an explanation and translation (see Figure 1) of a selected object of interest.

Locating items of interest

The student now recalls that he has accumulated similar articles of such a nature in the past and has probably kept some of these in his personal collection. There are a number of options for him to display the items in

Figure 3. 'Thesauri & Knowledge Package' in Object Explorer



Broader Term: Imagination
Vision

Focus Term: Visualisation

Narrower Term: 3D Test of Visual Skills
Visual Skills

Related Term: Imagery
Spatial Analysis

Used For: Data Mapping
Information Retrieval
Information Visualisation

Events: Seventh Int'l WWW Conference, 14 - 18 April 1998, Brisbane, Australia

People: Aha, D.W.; Belkin, N.J.; Card, S.; Cheng, Y.

Places: Norwegian Research Council; Pittsburgh University; University of Maryland

Things: Aggregate manipulator; Alpha slider; Animation; Attribute Explorer his personal collections. Figure 4 shows three different display options – a List view, a Map [17] view and a Forager [18] view.

The student can select the most suitable view for the task he wants to carry out. He can manipulate the information in the display to locate items of interest. He can, for instance, drag and drop items within the Forager view to bring different items into focus, and to get a quick view of the available items. He can also make use of pop-up labels and document detail panels to help identify articles of interest.

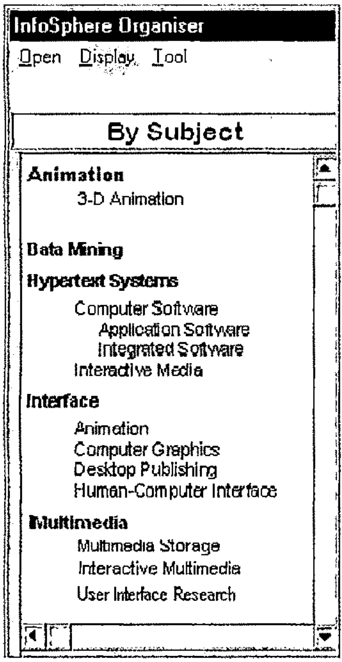
Extracting items of interest

Once the student has located an article of interest in *InfoSphere Organiser*, he can extract the article from this broader information world to his own personal workspace(s) within the PIE for further exploration. He highlights the selected article and drags it into *Object Viewer*. At this point of time, he can choose to close *InfoSphere Organiser*. It should be apparent that the integration of the broader information world and the student's personal workspaces has facilitated the launch of exploration based on current work, and hence, enables a uniform paradigm for accessing both his own personal workspace and the information world at large.

Browsing through items of interest

Once the selected article is dropped into *Object Viewer*, the first page of the article is displayed. The student can change the display to, for example, display the section headings of the article. He wants to quickly assess the document length, structure and layout. He does this quickly through the highlight-directed navigation – a co-ordination between *Object Viewer* and *Structure Viewer/Overviewer* workspaces. The tight co-ordination between these two workspaces provide a 'panorama view' – offering additional navigation aids by displaying an outline of the article in a separate navigation window through

Figure 4. Alternative views to display one's personal collections



Animation
- 3-D Animation

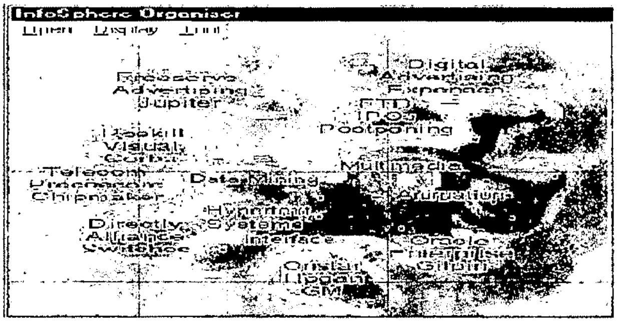
Data Mining

Hypertext Systems
- Computer Software
  - Application Software
  - Integrated Software
- Interactive Media

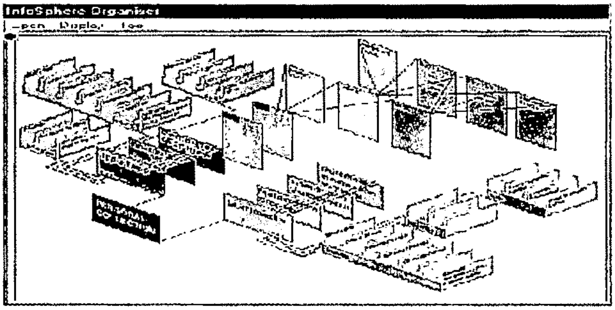
Interface
- Animation
- Computer Graphics
- Desktop Publishing
- Human-Computer Interface

Multimedia
- Multimedia Storage
- Interactive Multimedia
- User Interface Research

(i) List view



(ii) Map view



(iii) Forager view

which the student can move easily from page to page or section to section. By default, the 'Thumbnails' of the article pages are displayed in *Structure Viewer/Overviewer* (see Figure 1).

Other display options are available for the student's selection according to his preference and needs – such as a 'Book' [18] format or a 'DocLens' [19] format as shown in Figure 5. A 'Book format' allows the student to 'turn the pages' as if he is flipping through a printed document, bookmark the pages, and so on. A 'DocLens' provides an 'overview and detail' view. The student can use the 'outliner' to bring a different section into focus, and to examine different sections of the document.

Co-ordinated highlighting (navigation)

The student next wants to locate all the 'user interface' phrases within the article to assess the relevance of the article. Instead of scrolling through the whole article to do so, he decides to use the co-ordinated highlighting between *Object Viewer* and *Structure Viewer/ Overviewer*. He can also highlight more than one term or phrase and use different colours to distinguish them. By (colour-) highlighting a subset of information units that occur close together in a display, multiple display may be visually co-ordinated. Figure 6 shows an example where two terms have been highlighted using different co-ordinating colours and a 'Statistics' tool in the *Structure Viewer/Overviewer* is used to examine the distribution of the two terms within the article.

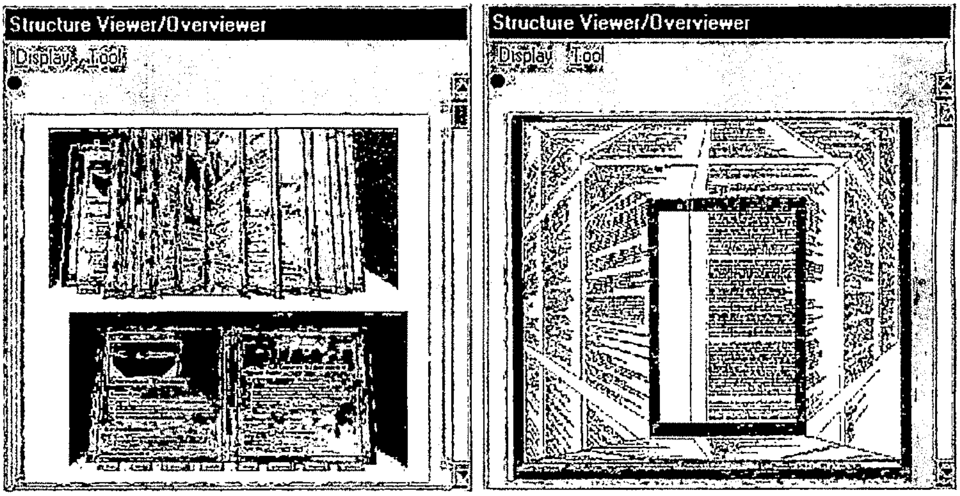
Filtering

While scanning through the article the student notes the pages containing the selected phrase and decides to have only these pages displayed so as to avoid cluttering *Structure Viewer/Overviewer* and to help him focus on these pages. He applies a 'Filter' from the 'Tool' menu in this workspace to achieve this.

'Focus+Context' viewing and exploration

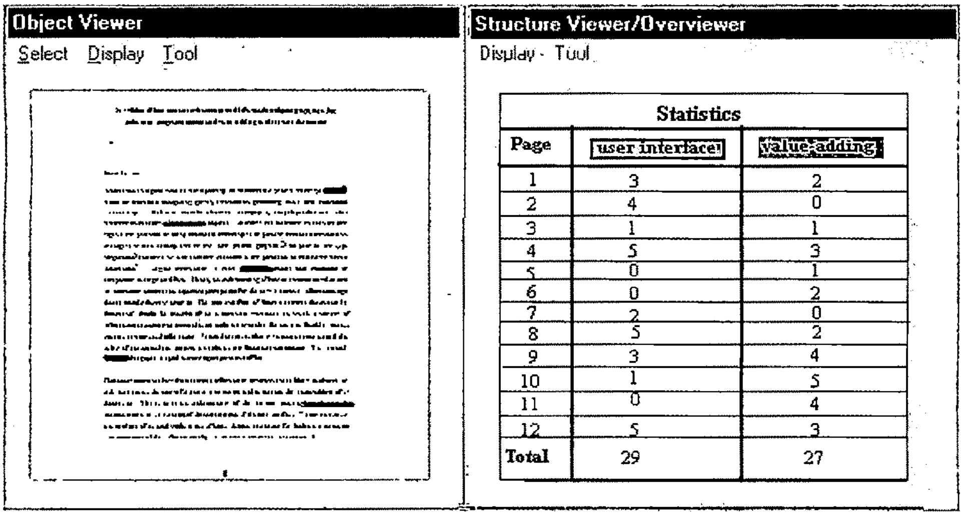
While browsing through the pages, the student comes across a link to another article that seems interesting. He decides to check out the details of the article but with the original article under exploration remaining in context. He highlights the link in *Object Viewer*, drags and drops it into *Object Explorer*. The 'body' of the link remains in its

Figure 5. Different visualisation modes to view a document layout/structure



(i) 'Book' format (ii) 'DocLens'

Figure 6. Co-ordinated highlighting used in determining terms distribution within an article



Statistics		
Page	user interface	value-adding
1	3	2
2	4	0
3	1	1
4	5	3
5	0	1
6	0	2
7	2	0
8	5	2
9	3	4
10	1	5
11	0	4
12	5	3
Total	29	27

original context while the 'clone' of the link is now in *Object Explorer* ready for further exploration. The student can also drag the same object twice into *Object Explorer* to simultaneously carry out two different operations on the same object (see Figure 7). The information-centric interface approach throughout the PIE allows for 'focus+context' view. The user can focus on different information objects in detail while still keeping them in the context of the full information set.

Controlling the amount of information

The student can control the amount of information (details) to be displayed by a zooming mechanism in the *Object Explorer*. Depending on the zoom level, different amounts of detail are displayed. At one level, only the abstract is displayed, for example. At another level, the abstract and breakdown of sections are displayed. Still at another level, the user gets the full-text view of the article. Zooming interface is an intuitive way to control the amounts of displayed detail smoothly in a graphical interaction environment.

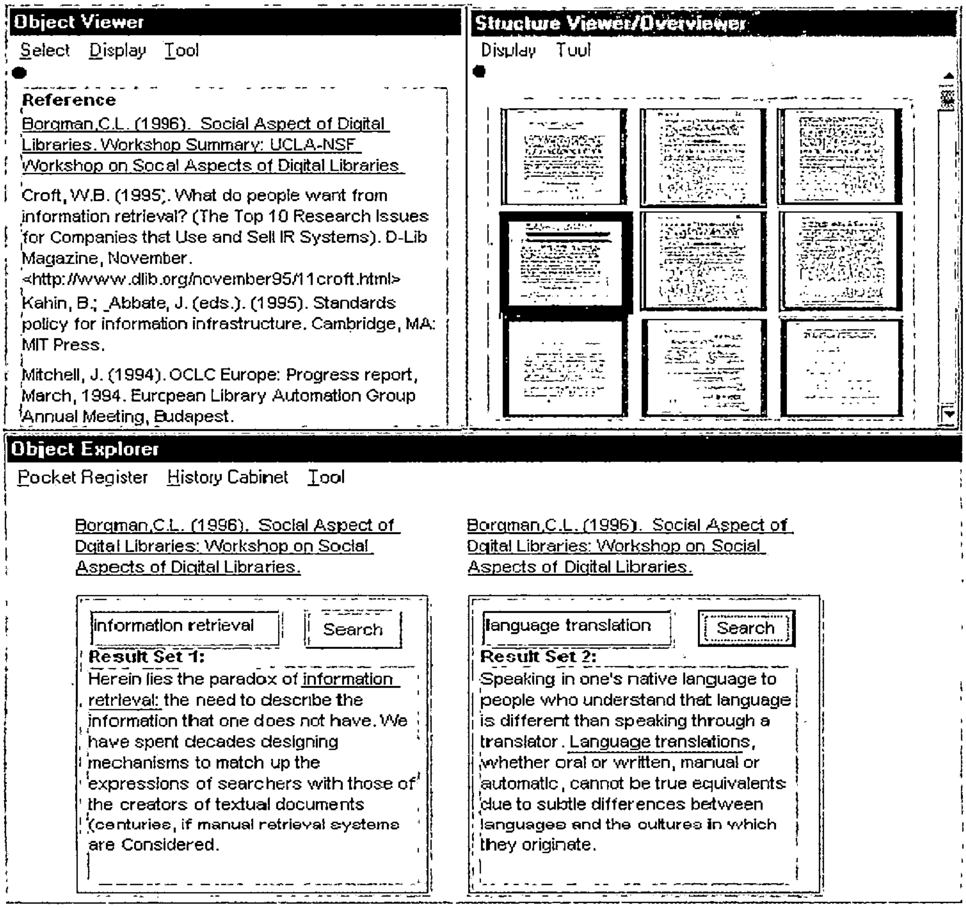
Collaborating

Tools are also incorporated within the PIE to allow the student to learn from other users and to contribute his thoughts and ideas. With the article still highlighted, the student selects the 'Comment- View' option from the 'Tool' menu in *Object Explorer*. He looks at what other similar researchers (including some renowned researchers in the field) have commented on the article. He can also add his comments (through the 'Comment- Add' option). Similarly, he can examine annotations by other researchers attached to the article through the 'Annotation- View' option. Annotations appear as thumbnail files embedded throughout the article at points where annotations have been added. The thumbnail images are displayed when the mouse cursor is paused over them. When there is only a single annotation embedded in the thumbnail, a double-click brings the student directly to the details (in *Object Explorer*). When there is more than one annotation, the user can browse and search through a list (database) of them. He can also add his annotation(s) to the article through the 'Annotation- Add' option.

Temporarily storing items of interest

At any time, should the student comes across an object of interest in any part of the environment, but wishes to explore it at a later

Figure 7. 'Body & Clone' and 'Focus+Context' view within the PIE



Reference

Borgman,C.L. (1996). Social Aspect of Digital Libraries. Workshop Summary: UCLA-NSF Workshop on Socal Aspects of Digital Libraries

Croft, W.B. (1995). What do people want from information retrieval? (The Top 10 Research Issues for Companies that Use and Sell IR Systems). D-Lib Magazine, November. <http://www.dlib.org/november95/11croft.html>

Kahin, B.; Abbate, J. (eds.). (1995). Standards policy for information infrastructure. Cambridge, MA: MIT Press.

Mitchell, J. (1994). OCLC Europe: Progress report, March, 1994. Eurcpean Library Automation Group Annual Meeting, Budapest.

Borgman,C.L. (1996). Social Aspect of Digital Libraries: Workshop on Social Aspects of Digital Libraries.

Result Set 1:

Herein lies the paradox of information retrieval: the need to describe the information that one does not have. We have spent decades designing mechanisms to match up the expressions of searchers with those of the creators of textual documents (centuries, if manual retrieval systems are Considered.

Borgman,C.L. (1996). Social Aspect of Digital Libraries: Workshop on Social Aspects of Digital Libraries.

Result Set 2:

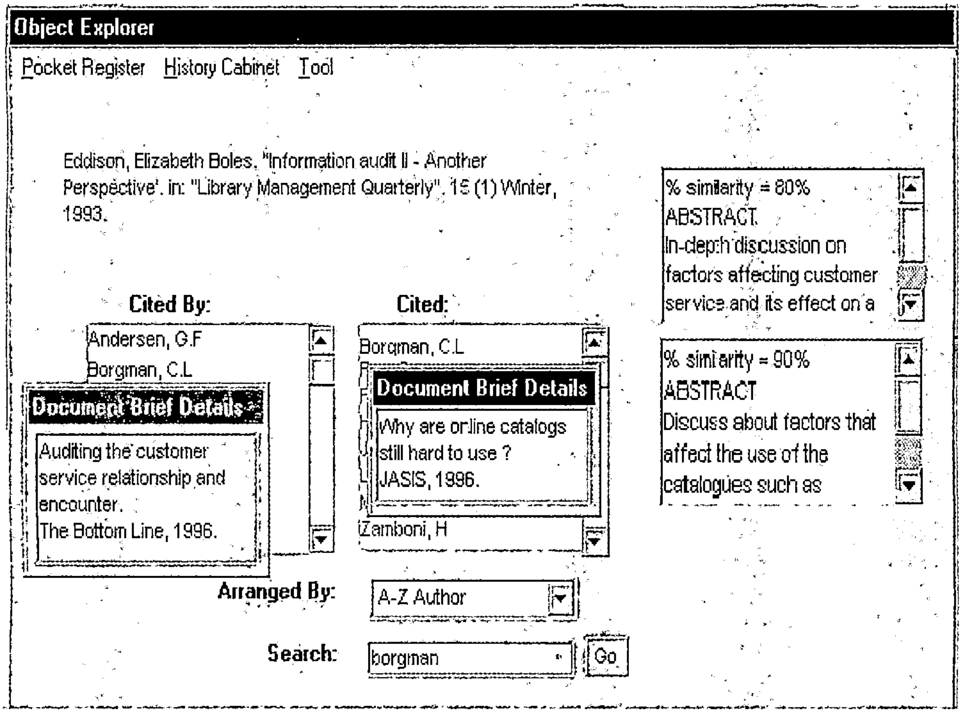
Speaking in one's native language to people who understand that language is different than speaking through a translator. Language translations, whether oral or written, manual or automatic, cannot be true equivalents due to subtle differences between languages and the cultures in which they originate.

stage, he can store it temporarily in the 'Pocket Register' available in *Object Explorer* while returning to his original task. The 'Pocket Register' hence, allows the student to carry out another task within the environment or return to the original task. All objects (e.g. articles, images, annotations) stored in the 'Pocket Register' are ready-at-hand for later exploration.

Browsing through related work

The student now wishes to examine other works that are related to this particular article. From the 'Tool' menu in *Object Explorer*, he selects the 'Research Compass' option and further chooses the 'Citation Linking' option. As shown in Figure 8, the 'Citation Linking' allows him to get a quick view of what (and how many) other works have cited this article and what other works this paper has cited (i.e. what previous works have influenced this study). He can manipulate the display to have the information arranged by author or publication date, or by percentage of similarity (i.e. degree of similarity of article selected to original article under exploration). He can also carry out a search to locate works based on specific author or keywords. The 'Citation Linking' applies the use of co-ordinated colours, pop-up labels and document detail

Figure 8. 'Citation Linking' – checking details using pop-up labels and panels



Eddison, Elizabeth Boles. "Information audit II - Another Perspective". in: "Library Management Quarterly". 15 (1) Winter, 1993.

Cited By:

Andersen, G.F

Borgman, C.L

Auditing the customer service relationship and encounter.
The Bottom Line, 1996.

Cited:

Borgman, C.L

Why are online catalogs still hard to use ?
JASIS, 1996.

Zamboni, H

% similarity = 80%
ABSTRACT
In-depth discussion on factors affecting customer service and its effect on a

% similarity = 90%
ABSTRACT
Discuss about factors that affect the use of the catalogues such as

Arranged By: A-Z Author

Search: borgman

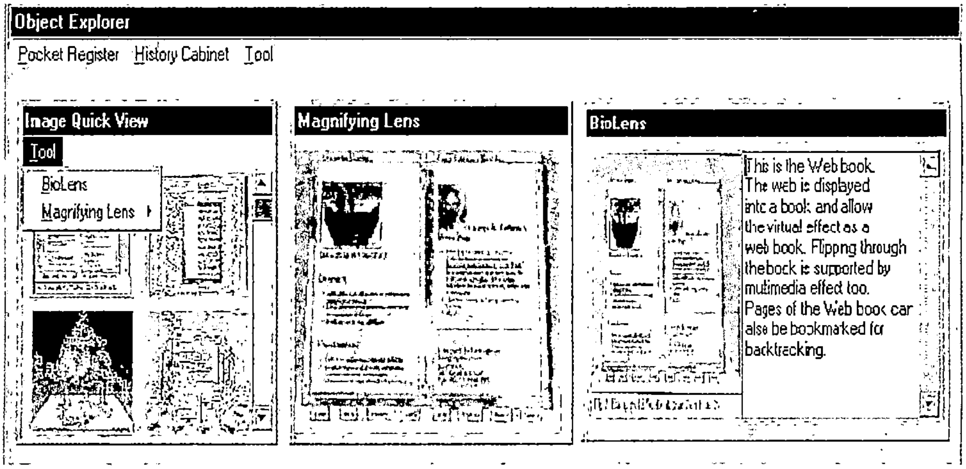
panels to present details-on-demand to the user (see Figure 8).

Interacting with multimedia objects

The student wants to view a live demo file embedded in the article without losing the context of the article. The demo is dragged into and run in *Object Explorer* while the article remains in *Object Viewer*. Next, he comes across a graph and some equations in the document. Instead of a static graph of an equation (with a particular parameter set to few common values), the equations and parameters are held here as a live-math set. The student inputs his own sets of data to the interactive equation provided. He interacts with parts of the author's data and observations by submitting a query to the author's programme – to, for example, produce graphs for values he chooses to enter (to, for instance, readily compare or verify the results presented).

The student wishes to get a quick view of all the images within the article. From the 'Tool' menu in *Object Explorer*, he selects the 'Image Quick View' option. A panel pops up, displaying a set of thumbnail images. Using the 'Magnifying Lens' in the 'Image Quick View' panel, the student zooms in (to magnify) at different levels for details of selected images. He uses the 'BioLens' to access detailed description and explanation of the images. An audio file that explains an image in auditory output can also be accessed. Figure 9 shows an example of a 'zoomed' image and the 'BioLens' in use. Within the 'Image Quick View' panel, the student can also arrange (and re-arrange) the images according to, for example, colour tendencies. For instance, to upward, images are aligned in the weight of a particular colour, to rightward, in order of the weight of another particular colour, to downward and so on.

Figure 9. 'Magnifying Lens' and 'BioLens' in use



This is the Web book. The web is displayed into a book and allow the virtual effect as a web book. Flipping through the book is supported by multimedia effect too. Pages of the Web book can also be bookmarked for backtracking.

Backtracking

The student decides that he has had enough exploration of the article. He now wishes to check out his temporary collection in the 'Pocket Register' (in *Object Explorer*). Using 'View Pocket', he scans through the list of items to determine whether he wants to explore them, or to save them into his personal collection in *InfoSphere Organiser* or to discard them. There is a search mechanism for a quick search and retrieval of items. After checking the details of each, he selects a couple of items to be saved into his personal collection. The integration of the broader information sphere and the student's personal workspaces has simplified this task. The student then realises that he is in need of some more images to illustrate the points he is trying to make for one of the papers he is working on. He decides to examine his current search history to find some potentially useful images. He does this by taking items stored in the 'History Cabinet' and uses them as query sets to locate relevant images.

Discussion

The above scenario shows how a user might interact with e-journals within the PIE. The PIE extensively attempts to satisfy the following six design requirements that have been derived from the set of desired properties as proposed in a previous paper [15]:

1. Integration of workspaces – The PIE integrates both the user's personal workspace (*Object Viewer*, *Structure Viewer/Overviewer* and *Object Explorer*) and the broader information world (*InfoSphere Organiser*) to allow easy transition between the user's main tasks and subtasks. This integration of the various workspaces also simplifies the task of consolidating intermediate and final seeking results into one's initial task.
2. User-controlled malleability – Throughout the PIE environment, users are able to perform direct manipulation on information objects.
3. Visual sovereignty – The user is provided with visual representation to the various conceptual items of the user's information gathering and analysis tasks. A collection of information objects related to the tasks is presented to the users. The student in the scenario has been given contextualised views during his exploration – for example, how an item fits within a collection set and within the broader information world.
4. Multiple representation and views – The PIE interface also allows the user to directly manipulate various aspects of the representation and view of information objects and their organisation. The user can

choose a view of interest from multiple display (and organisation) options according to his task(s) at hand and needs.

5. Integration with appropriate tools – The PIE provides users with a suite of interactive tools for querying, navigation, organising and authoring information. The user can, for instance, specify a query against a set of objects, browse and navigate with aids of visual clues about where items are, and seek for specific information by a zoom mechanism. The user can also make digital comments and annotations and attach these to an object (e.g. an article, a sentence, or an image).
6. Malleable information objects – To support a variety of information-seeking tasks, the information objects in the environment are fractionally structured. Users can manipulate the objects at multiple levels of granularity in a coherent manner. The structure of each object is also explicitly visible to the user – for instance, stating precisely where they fit into various larger structures. All objects at any granularity are queriable and navigable. We note how the student in the scenario can interact with a collection of articles, a selected article, a page of the article, and down to the level of interacting with an image or a word within the article. History sequences of the objects generated are visualisable and malleable (as allowed in the 'History Cabinet'). The information objects in the environment are also affiliated with a large prescription of metadata and other additional layers of information (e.g. comments, reviews and annotations attached) that are all value-adding properties that provide a user with additional information to facilitate decision-making.

Overall, we have observed the combination of characteristics arrived at in response to the design desideratum to be abounding – propagating a variety of possible rich interactions within one environment.

Current Work

The PIE has undergone a first-tier user evaluation. A scenario-based evaluation [20, 21] supported by 'concept prototyping'[1] is used to provide the platform and settings that are both realistic to real needs, and motivating and revealing for the evaluation. The scenario used is similar to the one presented in this paper. A set of 72 tasks with their respective interface mock-ups is used in the first-tier evaluation that involved 22 participants, most of whom are part-time graduate students undertaking the Masters of Science in Information Studies programme at Nanyang Technological University, Singapore. The evaluation was conducted in the 'Systems Analysis and User Interface Design' course as part of the hands-on laboratory sessions. The results and feedback gathered are currently being analysed and used as the basis to plan for the next tier of user evaluation. Certain aspects of the evaluation may be up-scaled or down-scaled if deemed necessary. This second-tier evaluation will be carried out with a larger pool of participants (targeted at between 50 to 60 participants). The set of 'features and tools' that is being tested in the first-tier evaluation is included in the Appendix of this paper.

This research seeks to further understand the potential and impact of an innovative information environment in enhancing user activities in using e-documents for various information-seeking purposes. Therefore, the overall objectives are to assess the representative users' overall perception and acceptance of the various concepts introduced within the PIE, and the strengths and weaknesses of the various features and tools in the interface.

Conclusion

In this paper, we have described the conceptual overview of an information environment proposed to support enhanced user-e-document interaction and also value-adding to such documents. Following the discussion of the overview of the environment, we present a set of scenarios to introduce and discuss in detail, the various features and tools within the PIE based on a set of information gathering and analysis tasks a user of e-journals performs. We demonstrate the use of the PIE throughout the scenarios, based on possible applications and utilizations of the environment interface in a real work setting. Current work on a first-tier evaluation of the PIE is briefly reported. The findings of these evaluations will be used to develop a set of recommendations to improve designs for enhanced interaction and value-adding to e-journals.

References

1. Kircz, J.G. Modularity: the next form of scientific information presentation? *Journal of Documentation*, 54(2), 1998, 210-235.

2. Kreitzberg, C.B. Designing the electronic book: human psychology and information structures for hypermedia. In: Salvendy, G. and Smith, M.J., eds. *Designing and using human-computer interfaces and knowledge-based systems*. Amsterdam: Elsevier Science, 1989, 457-464.

3. JISC. *Open Journal Project*. http://journals.ecs.soton.ac.uk/ (visited 9 September 1999).

4. *e-Lib Electronic Libraries Programme*. http://www.ukoln.ac.uk/services/elib/ (visited 9 September 1999).

5. JISC. *SuperJournal Project*. http://www.superjournal.ac.uk/sj/ (visited 9 September 1999).

6. Royal Society of Chemistry. *Chemical Communications*. http://www.rsc.org/is/journals/current/chemcomm/cccpub.htm (visited 10 September 1999).

7. *Earth Interactions*. http://EarthInteractions.org/ (visited 3 August 1999).

8. *Interactive Multimedia Journal for Computer-Enhanced Learning*. http://imej.wfu.edu/ (visited 10 September 1999).

9. Plaisant, C., Milash, B., Rose, A., Widoff, S. and Shneiderman, B. Life Lines: visualizing personal histories. *Proceedings ACM CHI*, 1996, 221-227.

10. Roth, S.F., Lucas, P., Senm, J.A., Gomberg, C.C., Burks, M.B., Stroffolino, P.J. and Kolojejchick, J.A. Visage: a user interface environment for exploring information. *Proceedings IEEE Information Visualization*, 1996, 3-12.

11. Masui, T. LensBar – visualization for browsing and filtering large lists of data. *Proceedings IEEE Information Visualization*, 1998, 113-120.

12. Bederson, B.B. and Hollan, J.D. Pad++: a zooming graphical user interface for exploring alternative interface physics. *Proceedings User Interfaces and Technology*, 1994, 17-27.

13. Mackinlay, J.D., Robertson, G.G. and Card, S.K. The perspective wall: detail and context smoothly integrated. *Proceedings ACM CHI*, 1991, 173-179.

14. Robertson, G.G. and Mackinlay, J.D. The document lens. *Proceedings ACM User Interface Software and Technology*, 1993, 101-108.

15. Liew, C.L. and Foo, S. Derivation of interaction environment and information object properties for enhanced integrated access and value-adding to electronic documents. *Aslib Proceedings*, 51(8), 1999, 256-268.

16. Gordon, A.S. and Domeshek, E.A. Déjà vu: a knowledge-rich interface for retrieval in digital libraries. *Proceedings Intelligent User Interface*, 1998, 127-134.

17. *NewsMaps – explore landscapes of information*. http://www.newsmaps.com/ (visited 10 September 1999).

18. Card, S.K., Robertson, G.G. and York, W. The WebBook and the Web Forager: an information workspace for the World-Wide Web. *Proceedings ACM CHI*, 1996, 111-117.

19. Robertson, G.G. and Mackinlay, J.D. The document lens. *Proceedings ACM User Interface Software and Technology*, 1993, 101-108.

20. Chin, G. and Rosson, M.B. Progressive design: staged evolution of scenarios in the design of a collaborative science learning environment. *Proceedings Computer-Human Interaction*, 1998, 611-618.

21. Carroll, J.M. ed. *Scenario-based design: envisioning work and technology in system development*. New York: John Wiley & Sons, 1995.

Note

1 'Concept prototyping' refers to a common approach in human-computer interaction (HCI) research in using prototyping packages such as MS Visual Basic, MS FrontPage, Powersoft PowerBuilder and Borland Delphi to simulate one's design, mockup or prototype. In this research, MS Visual Basic is used to simulate the various proposed design concepts, user interface mockups and interactions.

Appendix –'Features/Tools' currently tested in each part of the PIE workspaces.

Workspace/ 'Features/Tools'		**Description**
InfoSphere Organiser		
Open		User uses this menu to access either his personal collection or external collections.
Display		
(Pop-up labels and panels are employed within these various displays to present details-on-demand on user request)	CamTree	A display of collections/documents in a 'tree' form – enabling a quick view of collections or items within the collections. User can easily 'rotate' the 'tree' to bring different items into focus.
	DocTOC	Depicts the overall contents of the items in the collection (e.g. each article) and showing the size and types of media in each item (e.g. size in number of bytes of text/images/ audio/video).
	Forager	Provides the user with a panorama view of collections. User can drag and pull items of interest to the front of the display (see Figure 4).
	List	Lists and sorts items according to various entities such as author, title, publication date. User can scroll along the lists to locate items of interest, or employ a search mechanism.
	Map	Shows the collection in the form of a 'topographical map' of mountains and valleys. The concept of the layout is that documents with similar contents are placed close together, and 'peaks' emerge where there is a high concentration of documents. Documents containing the same thematic terms appear roughly at the same point (see Figure 4).
	Time Line	Used to depict the distribution of works done (e.g. by topics) using a graph form according to a time line. User can select each point in the graph to obtain additional information of the selected point (see Figure 2).
Tool	Zooming	User can use this tool to zoom for more details (-on-demand) of selected items of interest from the collections of documents, or to get a magnified view of a selected item.
Object Viewer		
Select		User can select various information objects in the document (e.g. a whole article, a highlighted paragraph, a highlighted page, or a highlighted term, diagram, word, etc.) for further exploration.
Display		Displays the content of a document according to user-defined preference (e.g. page-by-page, selected section headings, etc.).
Tool	Find/Search	Find/search feature is used for locating specific information/section (e.g. specific term/word/image) within the article. The found items are automatically highlighted for the user's attention.
	Highlighter	To highlight item(s) of interests. Different colours/highlighting methods (e.g. underlining) can be used to distinguish various items. Used for 'coordinated-highlighting/navigation' purposes (see Figure 6).
Structure Viewer/ Overviewer		
Display	Book	Display the document in a book format (see Figure 5).
	DocLens	Display the document using a Document Lens (see Figure 5).

	Thumbnails	Display the document as a series of thumbnail images (see Figure 1).
Tool	Arrange	A tool to arrange information for display. User can juxtapose or superimpose items, or freely arrange items within the workspace, such as placing focussed item in the centre of attention, for quick referencing and cross-referencing, etc.
	Filter	A tool to help reduce space contention by filtering out unnecessary or irrelevant details on display to help focus on more important information.
	Statistics	To check distribution of selected information objects (e.g. terms/phrases) to assess a document's relevance, for instance. Provides a summary/ quick view of distribution (See Figure 6).
	Zooming	A zooming tool for more details (-on-demand) of selected items of interest from a document, or a magnified view of a selected item (e.g. an image).
Object Explorer		
Pocket Register		Serves as an interactive, multi-object, visible 'back pocket' that allows the user to temporarily store items of interest for later exploration. (Also refer to scenario discussion on 'Temporarily storing items of interest' and 'Backtracking').
History Cabinet		Stores user's search and browse activities. The historical records are malleable for the purpose of, for example, Backtracking.
Tool	Alerting/ Updating	A tool to inform the user of recent works related to a selected item/chosen topic. A number of different displays (e.g. list, time line) can be selected. The support of an intelligent agent that helps define 'clusters' of related work is assumed here.
	Annotation-View/Add	A tool for viewing and adding annotations in the document (Refer to the scenario on 'Collaborating').
	Comments-View/Add	A tool for viewing and adding comments about the document (Refer to the scenario on 'Collaborating').
	Glossary/ Definition Lens	A tool to present user with definitions/ glossaries related to an unfamiliar information object (e.g. word/term).
	Image Quick View	Tools to show and manipulate a list of thumbnails of all images in the document (see Figure 9).
	Research Compass	For browsing related work using one of the following 3 tools:
	Citation Linking	Shows all 'cited' and 'cited by' works of the document (see Figure 8).
	Document Mapping	Mapping item of interests using a 'topographical map' concept. User can choose to map by various options (e.g. topics, author). Pop-up labels and panels are used to display details-on-demand.
	Document Similarity	Adopts a 'galaxy' concept that clusters related work according to four closest topics/criterion (based on the content of the item of interest) and uses a variety of visual coding principles, e.g. shape/colour coding to indicate the number of criteria satisfied; proximity coding whereby the closer a document is located to a criterion, the more likely it is that the document contents are related to it.

Thesauri & Knowledge Package	A thesauri tool plus a list of related events, people, places, things that are associated with the selected information object (see Figure 3).
Translation	A translation tool to support selected term/word/article translation into various languages (by selection) (See Figure 1).
Zooming	Zooming tool for more details (-on-demand) of selected items of interest, or magnified view of a selected item (e.g. an image).

In the catalogue ye go for men: evaluation criteria for information retrieval systems

Julian Warner
School of Management, The Queen's University of Belfast, Belfast BT7 1NN
j.warner@qub.ac.uk

The contrast between the value placed on discriminatory power in discussions of indexing and classification and on the transformation of a query into a set of relevant records dominant in information retrieval research has not yet been fully explored. The value of delivering relevant records in response to a query has been assumed by information retrieval research paradigms otherwise differentiated (the cognitive and the physical). Subsidiary concepts and measures (relevance and precision and recall) have been increasingly subjected to critiques. The founding assumption of the value of delivering relevant records now needs to be questioned. An enhanced capacity for informed choice is advocated as an alternative principle for system evaluation and design. This broadly corresponds to: the exploratory capability discussed in recent information retrieval research; the value of discriminatory power in classification and indexing; Giambattista Vico's critique of the unproductivity of Aristotelian methods of categorisation as routes to new knowledge; and, most significantly, to ordinary discourse conceptions of the value of information retrieval systems. The criterion of enhanced choice has a liberating effect, restoring man as an artificer and enabling a continuing dialectic between theory and practice. Techniques developed in classic information retrieval research can be adapted to the new purpose. Finally, the substitution of the principle of enhanced choice exemplifies the development of a true science, in which previous paradigms are absorbed into new as special cases.

Ay, in the catalogue ye go for men;
As hounds, and greyhounds, mongrels, spaniels, curs,
Shoughs, water-rugs, and demi-wolves, are clept
All by the name of dogs: the valu'd file
Distinguishes the swift, the slow, the subtle,
The housekeeper, the hunter, every one
According to the gift which bounteous Nature
Hath in him clos'd; whereby he does receive
Particular addition, from the bill
That writes them all alike;
Shakespeare. *Macbeth.* c.1606.

Introduction

The epigraph indicates the value that has been historically attached to subtlety of distinctions in the language or lexicon of information retrieval systems. In this respect, the passage anticipates the principle formulated in modern discussions of indexing and classification that the value of an index term lies in its discriminatory power. In this principle, and in its historical anticipation, there is a strong, although largely unnoticed, contrast with the assumption of information retrieval research, particularly experimental information retrieval research, that the performance of an information retrieval system is to be measured by its capacity to deliver relevant records in response to deliberately articulated queries.

The concern here is, then, with whether the central principle embodied in the practice and theory of classification and indexing can yield more satisfying design and evaluative criteria for information retrieval systems than those which have been characteristically assumed in information retrieval research. Two paradigms have been distinguished in information retrieval research, the cognitive and the physical, but they share the assumption of the value of delivering relevant records [1, p. 19; 2]. For the purposes of the discussion here, they can be considered as a single, if heterogeneous, paradigm, linked if not united, by this common assumption.

The contrasting paradigm implicitly embodied in classification and indexing may finally be incommensurable with that of information retrieval research, with disputes not logically resolvable within either paradigm. The approach taken in this paper will be to suggest:

- that an alternative principle involving discriminatory power has been held, implicitly and explicitly, in a number of largely separate discourses;
- that the cumulative effect of recognising this is to indicate more viable and productive criteria for designing, using and evaluating information retrieval systems;
- and, finally, that the classical tradition of information retrieval research can itself be assimilated to the new model.

In this final respect, the development proposed here is an exemplar of scientific development in which discarded paradigms are absorbed into developing ones, as special cases.

The discourses in which an alternative principle for the design and evaluation of information retrieval systems can be discovered and which are to be covered here are:

- an emerging, although rather isolated and discontinuous, strand of information retrieval research;
- discussions of the accepted principles of classification and indexing;
- Giambattista Vico's critique of Aristotelian principles and categories for classification;
- and, most crucially, ordinary language discussion of information systems.

The paper will review information retrieval research, taking the liberty of conflating distinguishable aspects for the purposes of a higher level discussion, and then indicate that an alternative principle for evaluation can be found in the discourses identified. The value of the alternative model developed will be explored. In conclusion, it will be suggested that the alternative principle and criteria developed can have a liberating effect, allowing theory and practice to interact, and that a productive transformation of the field has been indicated.

Alternative paradigms

Information retrieval research

Information retrieval research, particularly in the experimental tradition emerging in the 1950s in Britain and North America, has taken as its founding assumption the principle that an ideal information system should deliver all (and possibly only all) the records relevant to a given information need. In order to evaluate information systems in relation to this desired end, or variations on it, various steps were taken: relevance was stabilised and quantified, sometimes being reduced to a binary or dichotomous variable; and measures of precision and recall, which depend on the prior stabilisation of relevance, were developed. More recent research has questioned the validity of aspects of this paradigm, although more frequently with reference to its subsidiary concepts (relevance and information need) and derived measures (precision and recall) than with regard to its founding assumption.

The adequacy of the concept of relevance employed in information retrieval research has been questioned. Experiments substitute a measurable phenomenon, relevance as constructed under artificial conditions, for an unmeasurable one, relevance under operational conditions, but fail to demonstrate that there is an adequate correlation between the two. Most disturbingly, it has been suggested that operational relevance is fluid, influenced by intention, context, and other documents seen or read, and simply not amenable to stabilisation or, further, quantification [1, 3].

The classical measures of precision and recall are also rendered increasingly artificial by the high degree of interactivity enabled by recent information technology developments. How, when searching a CD-Rom database, is the final set of records to be isolated except by a process whose very arbitrariness invalidates it as a component of a measure of system performance [4]? High interactivity, and unmediated searching, also reduce the need for a query to be fully articulated in advance of searching. There has also been a realisation that a deliberately stated query (which can be distinguished from an information need or assertion of relevance) may be a methodological requirement for controlled experiment, but is not intrinsic to the information seeking situation [5] and that it is possible to search without verbalising an information need.

The classic information retrieval paradigm, and the concepts and measures associated with it, could be preserved but only at the cost of increasing its distance from more realistic information seeking situations. It may be that not only are the classical concepts and measures both becoming and being recognised as increasingly artificial but that the founding assumption – that a system should deliver all (and only all) the relevant records should be re-examined. What is required, then, is not questioning of concepts with the paradigm but of its founding assumption, turning what has been received as a given into an object of enquiry.

To some extent, this has begun to occur within information retrieval research. The subtlety and complexity of information retrieval has been recognised [6]. Most specifically, the principle of exploratory capability, the ability to explore and make discriminations between representations of objects, has been suggested as the fundamental design principle for information retrieval systems [1, 3].

On a subjective level, this can be supported by introspection: that what I desire from an information retrieval system is not a possibly mysterious transformation of a query into a set of records, but a means of enlarging my capacity for informed choice between the representation of objects within the given universe of discourse. Such an enhanced capacity for informed choice broadly corresponds to exploratory capability. It could also be regarded as analogous to a sense of cognitive control over, or ability to discriminate between, representations of objects.

One example, which may be fictional in a double sense, can be given of the need for enhanced discriminatory power. At one point in time, a researcher might wish to distinguish the private individual, Samuel Langhorne Clemens, from the author, Mark Twain (perhaps out of interest in his copyright disputes or in his brother's, Orion Clemens, activities as Secretary of Nevada Territory). What would be valuable for this purpose would be a system that did not conflate these two distinguishable aspects of the individual but enabled them to be differentiated. At a later point in time, the same researcher might be interested in information on Mark Twain and Samuel Clemens considered as a single entity. An information retrieval system should then be able to differentiate and to link together the occurrences of these different names, as required.

In conclusion, the assumption that it is desirable to obtain all, and possibly only all, the records relevant to a given query can be rejected in favour of the alternative principle of exploratory capability or enhanced capacity for informed choice. Introspection supported the value of exploratory capability. Its appeal as an alternative to the established information retrieval paradigm could be strengthened if it could be found, even if only implicitly or in analogous forms, in other, independently developed, discourses.

Principle of indexing and classification

An acknowledged principle of indexing and classification is that the value of a term is its discriminatory power. By discriminatory power is understood the ability to partition and select from the objects represented within the given universe of discourse. What particular terms or methods of classification are appropriate will vary with the area of discourse and the focus of interest: most obviously, a factor that differentiates one set of objects from another will not serve to discriminate within either set of objects. Discriminatory power is again analogous to exploratory capability, or, more accurately, a critical factor in enabling progressive and controlled exploration.

Vico's critique of Aristotelian classification

A strong, and highly significant, analogue to exploratory capability can be found in Vico's critique of Aristotle. Aristotle's philosophy, as well as being a direct and indirect source for subsequent understandings of genus, species, specific difference, synonymy and equivalence, involved, in some of its aspects, a systematic method of enquiry in order to classify an object. An enquirer was required to ask a series of questions: Does the thing exist? What is it? How big is it? What is its quality? and the like. This method of enquiry was subjected to an incisive critique by Vico:

> Aristotle's *Categories* and *Topics* are completely useless if one wants to find something new in them. One turns out to be a Llull or Kircher and becomes like a man who knows the alphabet, but cannot arrange the letters to read the great book of nature. But if these tools were considered the indices and ABC's of inquiries about our problem [of certain knowledge] so that we might have it fully surveyed, nothing would be more fertile for research [7].

The last clause of that critique deserves emphasis: 'nothing would be more fertile for research'. The rigidity of the method is avoided, while some of its techniques are retained, and it is transformed into a systematic and effective means for enhancing knowledge of an object. Analogously, while rejecting the rigid transformation of a query into a set of records assumed as desirable in information retrieval research, similar techniques can be used to explore the domain of discourse covered by the information retrieval system.

A further supporting analogue can be found in the fictional rather than discursive treatment of rigid classifications in Dickens' *Hard Times*. The logical distinctions exemplified in Bitzer's definition of a horse – 'Quadruped. Graminivorous. Forty teeth, namely twenty-four grinders, four eye-teeth, and twelve incisive ... Age known by marks in mouth.' [8, p. 6] – which does resemble 19th century taxonomies for the horse, themselves influenced by the Aristotelian method of definition by genus and species, are presented as harsh [9, pp. 106-108]. Outside the restricting enclosure of the town, a different metaphor for knowledge is discernible:

> They walked on across the fields and down the shady lanes, sometimes getting over a fragment of a fence so rotten that it dropped at a touch of the foot, sometimes passing near a wreck of bricks and beams overgrown with grass, marking the site of deserted works. They followed paths and tracks, however slight [8, p. 353].

The value of an information system could then be the ability it offers discriminatingly to follow 'paths and tracks, however slight'. Classification schemes themselves (and their analogues in thesaural relations among indexing terms) can then be received not as fixed models of stable entities but as valuable exploratory devices.

Ordinary discourse

Ordinary, particularly informal spoken, discussion of information systems is simultaneously highly significant and difficult to produce as evidence. Evaluative criteria may be implied rather than explicitly articulated. Yet when a searcher complains that it is difficult to control the number of records retrieved, a principle of discriminatory power is being invoked. More explicitly, one spoken response to an earlier version of this paper was: 'that's the basis [an enhanced capacity for informed choice] on which people use systems anyway'.

Summary

Similarities in themes and principles enunciated or implied have been revealed in largely separate discourses, emerging in information retrieval research, implied in discussions of principles of indexing and classification, made explicit in Vico's critique of Aristotelian methods of investigation, and present, in partly unarticulated form, in ordinary discourse. The mode of expression varies, but an enhanced capacity for informed choice, for effective discrimination, or for cognitive control was discovered to be valued in all the discourses adduced. Independent agreement with an emerging and rather isolated theme in information retrieval research, of exploratory capability, offers support for replacing the established emphasis on the delivery of relevant records with such a principle for the design and use of information systems. In some respects, possibly through the influence of practices of indexing and classification and of ordinary discourse understandings, working systems may offer exploratory capability and productive interaction. In Vico's terms, practical understanding has been in advance of theoretical articulation.

Evaluative model

Endorsing the principle of enhanced capacity for informed choice can have a liberating effect, revealing the intra-theoretic nature of many disputes within the classic tradition of information retrieval research: it offers the possibility of a deeper understanding of relevance; enables a mutually informing relation between practice and theory; restores man as artificer as a user of information systems rather the cipher of information retrieval research; and can enable the development of more satisfying evaluative criteria.

Disputes over the validity of constructs demanded for retrieval system evaluation in the classic tradition of information retrieval research, for instance whether deliberately contrived relevance judgements are adequately correlative with real world judgements, can now be regarded as intra-theoretic, connected with the theoretical framework imposed, not inherent in the process of information retrieval and not necessarily contributing to an understanding of those processes. In some respects, the construction imposed by the research paradigm may even have inhibited development of understanding of its chosen domain of study. For instance, the methodological need to reduced relevance to assessments, possibly open to quantification, and stable over time, may have inhibited exploration of its many possible dimensions. Some dissenting discussions have insisted on its complex and multi-faceted nature [10, 11].

A mutually informing and productive relation between theory and practice can be developed. For instance, the practical experience of those indexing procedures or retrieval algorithms which enhance exploratory capability in specified circumstances can inform theoretical development and system design and modification. The divorce of information retrieval research from practice has been noted and sometimes regretted, although less often explained. Now the practical understanding embodied in working systems can be recognised and theoretically developed.

The further question then also arises as to whether accepting the principle of exploratory capability has practical implications in terms of the indexing procedures or algorithms for searching to be adopted. An immediate response would be that it does not necessarily have unambiguous practical implications: that the particular indexing procedures and algorithms to be used will be critically dependent on the purpose and context of retrieval, including the cost of indexing and retrieval. Crucially for continuity of systems development, techniques identical with, or analogous to, those currently developed may be used to different ends. It should also be noted that the Boolean logic used in many retrieval systems, does, under certain conditions, have the advantage of relative transparency to the searcher. The objection that it is an ineffective way of transforming an information need into a set of relevant records is no longer tenable. It could still be objected that in some applications, for instance with heterogeneous written material without humanly assigned index terms, it gives inadequate control over the representations within the universe of discourse.

A deeper effect is to restore man as an artificer and to recognise the subtlety of the processes involved in information retrieval. Rather than being subjected to a retrieval process beyond immediate control, the searcher is presented with an enhanced capacity for choice and for making recalled sets. The new, and historically unprecedented, potential for enhanced forms of knowing of existing textual material can then be productively explored. For instance, the unrivalled opportunity offered by full text database for exploring the semantic mutability of written word forms with different contexts can be pursued.

More detailed evaluative criteria could be developed from the central evaluative principle of enhanced choice, partly by drawing on the understandings already developed in discussions of classification. Yet it should be recognised that quantitative comparative measures are unlikely to result. Once the diversity of contexts for information retrieval is recognised, the idea of a single, generally applicable approach to system design, or a single comparative measure of system performance, becomes severely questionable. The best outcome which can reasonably be expected from research and from reflection on practice is a better understanding of the process of information retrieval, which can then be used either to design better information systems or to make more effective use of existing systems.

Conclusion

Replacing the emphasis on the delivery of relevant records with a stress on exploratory capability or cognitive control as a design and use principle for information retrieval can have a liberating effect. It yields more satisfying evaluative criteria while preserving a strong continuity with previous work, particularly in recognising the utility of developed information retrieval techniques. Theory and practice, rather than being separate or even antagonistic, are enabled to inform each other. The discourses, of philosophy, classification and ordinary discussion, from which the new principle has been drawn, can be brought further to bear upon information retrieval, transforming it into a human science and recognising its subtlety and significance. A minor, although significant, relief is liberation from the obligation to read work in the classic information retrieval paradigm, except for the emerging signs of self-questioning.

The transformation advocated in this paper resembles, in some respects, a mathematical revolution and can also be seen as an example of scientific progress. Classically fundamental transformations of mathematics have preserved the form while modifying the interpretation [12]; analogously, information retrieval techniques have been preserved but adapted to a new end. More broadly, it has been argued that a discipline exhibits the history of a true science if its earlier stages can be seen as special cases, from the perspective of its subsequent development [13]: in this context, the automatic transformation of a query into a set of records can be seen as a possible support for informed choice, valuable in certain sets of circumstances. Rather than, as Swanson [6] indicated: 'Waiting for Godot [while failing] to grasp what is now within reach', we can begin to explore the potential for improving human interaction with recorded knowledge.

References

1. Ellis, D. Theory and explanation in information retrieval research. *Journal of Information Science,* 8, 1984, 25-38.

2. Belkin. N.J. and Vickery, A. *Interaction in information systems: a review of research from document retrieval to knowledge-based systems* (Library and Information Research Report 35). London: British Library, 1985, 114.

3. Ellis, D. *Progress and problems in information retrieval.* London: Library Association, 1996.

4. Warner, J. Retrieval performance tests in relation to online bibliographic searching. In: Shaw, D., ed. *ASIS '92: Celebrating change: information management on the move.* (Proceedings of the 55th ASIS Annual Meeting, Pittsburgh, PA, 26-29 October 1992.) Medford, NJ: Learned Information for the American Society for Information Science, 1992, 231-241.

5. Heine, M.H. The 'question' as a fundamental variable in information science. In: Harbo, O. and Kajberg, L., eds. *Theory and application of information research.* (Proceedings of the Second International Research Forum on Information Science, 3-6 August 1977, Royal School of Librarianship, Copenhagen.) London: Mansell, 1980, 137-145.

6. Swanson, D.R. Historical note: information retrieval and the future of an illusion. *Journal of the American Society for Information Science,* 39(2), 1988, 92-98.

Teaching infopreneurship: students' perspectives

Adeline du Toit
Rand Afrikaans University, PO Box 524, Auckland Park 200, South Africa
asadt@LW.rau.ac.za

A university course in infopreneurship could contribute to bridging the gap between the so-called security of full-time employment and the imagined insecurity of an own information business. At present the Rand Afrikaans University in South Africa offers infopreneurship courses to Information Science and Computer Science students. The course content and presentation methods of the courses are similar in nature, but the same cannot be said about the two groups of students. It was therefore decided to conduct a questionnaire-based survey of the two groups of students and to compare their responses. The exploratory investigation reported here scrutinisés the status quo of infopreneurship education at universities. The perceptions of the two groups of students with regard to the objectives of an infopreneurship course at the university, course contents and presentation methods are compared. In conclusion possible adaptions and improvements to the course are considered.

Introduction

Entrepreneurs are those individuals in society who take the lead as well as the risk in mobilising the production factors (natural resources, human resources and capital) in specific combinations to produce products and services for their communities. Those parts of the world that foster an entrepreneurial culture are characterised by economic wealth – something that is necessary for social and political stability. This is the case in countries such as Japan, those in Europe and the USA. The opposite occurs in countries that lack entrepreneurial economies, such as those in Eastern Europe and Africa. Presently, there are signs of a developing entrepreneurial culture in South Africa. Research by Lascaris and Hunt in 1998 showed that 69% of young South Africans would rather work for themselves than work in an existing enterprise [1]. A reason for this may be that the world economy is at present characterised by the phenomenon of shrinking job opportunities experienced in most countries. This is also the case in the information industry in South Africa.

The smaller demand for graduates at universities in South Africa, as well as the low economic growth rate and the serious unemployment problem, complicates the task of information specialists finding employment at existing information services. This state of affairs will force more information science graduates to start their own businesses. However, to expect people, who studied for the sole reason of securing a fixed appointment one day, to take the risk of starting their own businesses could be unrealistic. In this regard a university course in infopreneurship could contribute substantially to bridging the uncertainty gap between the so-called security of steady employment and the imagined insecurity of an own business.

In this article a literature review of developments in infopreneurship are discussed. Based on the findings of the literature review, a questionnaire focusing on aspects such as the objectives of an infopreneurship course at a university, course content and presentation methods was developed. A questionnaire based survey was then conducted among two groups of students in the Department of Information Studies at the Rand Afrikaans University, Johannesburg, South Africa. This was specifically done to gauge the opinion of students as customers. The respective responses of the two groups were compared and recommendations made on how to improve the teaching of infopreneurship at universities.

Infopreneurship as a subject

Information, the essential ingredient in today's knowledge-based economy, is dynamic, expensive, and continually being reconfigured and repackaged [2]. Enterprises are prepared to pay for information that will enable them to increase their profit margins and remain ahead of their competitors. A wide variety of enterprises and individuals have seized the opportunity of turning information provision into a successful commercial venture. Although a few of these ventures have issued from libraries, the majority have originated from the business sector. Many of these businesses provide the same type of information provided by libraries, whereas others provide information which libraries have either ignored in the past, or which they have regarded as being of no real value to anyone. Effectively managing organisational information and knowledge provides information professionals with the key to personal and professional competitive advantage.

Today, information specialists are compelled to use new skills and strategies in order to change, survive, and continue to compete in the world of virtual information. Weitzen [3] identified six ways in which infopreneurs generate, organise and disseminate information profitably. These are broadly accepted by different researchers as the way to operate infopreneurship. The six ways identified by Weitzen are:

- **Leveraged information:** Infopreneurs make use of databases or the Internet to obtain general market information and support data for research, as well as for the creation of marketing plans, pricing strategies, distribution channels and advertising programmes.
- **Customised information:** Information that is synthesised and tailored to the user's needs. This included the manipulation of information into a format that is easily understandable (e.g. computer generated graphics).
- **Facilitating access to information:** The information explosion of the last few years increased the need to find the right information in a timely manner. The gap between information produced and information consumed continues to widen, because of the rapid developments taking place today. One solution to this problem is electronic access to information, because this allows infopreneurs to access desired information, select what is valuable and, finally reconstruct and disseminate information selectively to clients.
- **Speeding up the flow of information:** The time-sensitive nature of information requirements ensures a high demand for the speeding up of the flow of information. Today infopreneurs are capable of generating, sending and receiving data virtually instantaneously via the World Wide Web, from and to remote locations thousands of kilometres away.
- **Repackaging information:** The process of repackaging information includes the organising of information into new information products and services. The goal of repackaging information is to locate unrelated and yet possibly relatable information and create new products combining the information.
- **Around the clock delivery of information:** As computers begin to substitute human effort, it becomes possible to deliver information instantly at any time of day or night. This process causes information to take on a value-added quality.

Even though entrepreneurship is widespread in the world today, it is still unclear what exactly makes some people good entrepreneurs. MaCleod [4] found that the qualities that are shared by successful entrepreneurs include a need for achievement and acknowledgement, a desire to use their skills to the full, and to be in control of their environment. Above all, entrepreneurs also tend to be those people who are willing to take risks.

The past thirty years have seen an enormous increase in the number of universities that started to offer entrepreneurship as a subject. Zeihaml and Rice [5] found that there was a high demand for entrepreneurship courses at both undergraduate and graduate level at the majority of the 140 universities surveyed in the USA. In South Africa, the majority of universities are now offering entrepreneurship as a subject or planning to introduce it in the near future. It has become crucial that departments teaching Information Science in South Africa include infopreneurship courses in their curricula.

During the seventies, information brokering and consulting services experienced a period of expansion, specifically in the USA. The industry became well developed in the USA, with courses being conducted and papers delivered at conferences by members of the industry. In Europe a code of conduct was developed for information brokers and information consultants [6]. In South Africa there was also a marked growth in demand for the services of information consultants. This was strongly stimulated, for example, by the need to manage information as an economic resource in enterprises. While there appeared to be a wealth of IT equipment in many enterprises, information skills were lacking in staff who were expected to be able to control their own systems, and to search cost-effectively and knowledgeably for information in internal and external systems. In a survey by Kotze and Boon [6], a profile of the South African information broker and consultant revealed the following:

- **General characteristics:** The industry is relatively young and not established and designations are used loosely. Practices are usually registered as closed corporations. Vocational experience is regarded as an important prerequisite. Most information brokers and consultants enter the industry after identifying a specific need in the market.
- **Resources:** Most information brokers and consultants practise from their homes. The industry is not capital intensive. Information brokers and consultants usually employ part-time personnel.
- **Services:** A large variety of information services and products are rendered. These services/products vary from system analysis and design to the analysis of information needs and the evaluation of software. Information brokers generally deliver information retrieval services and research services.
- **Market:** Clients, mostly persons with professional qualifications, are usually drawn from the private sector. The majority of clients come from the Pretoria-Witwatersrand area. Services are also rendered to overseas clients. Most clients repeatedly use the same information broker or consultant.

In general, South African information brokers and consultants were positive and optimistic about the future of their industry in South Africa. They indicated that success depends on the following conditions:

- clients should be made more information conscious and regard information as an essential resource;
- information brokers and consultants should adjust to changing needs;
- good market research should be undertaken;
- value-added information services and products should be rendered.

Infopreneurship is action-orientated. The design of a course on infopreneurship should encourage imagination, creativity and innovation, and address the issues of setting a balance between teaching entrepreneurial skills and managerial skills. Entrepreneurs identify needs in changing environments; they create and innovate with active imagination, a quality of a developed right hemisphere of the brain. Traditional schooling focuses on developing the left hemisphere of the brain. Entrepreneurship education should address the need to train students in the skills and knowledge required to bring an idea, invention, product or process from the stage of conception to its full commercial potential [7].

Entrepreneurship education should place greater emphasis on practical experience and application. Since entrepreneurs are action-orientated rather than concept-motivated, the art of learning skills should be developed by increasing awareness. Pragmatic demands for relevance and the application of knowledge will be required, while the student has to fully integrate the scholarly and practical learning styles, therefore making the lecturer's role that of a facilitator of the learning process, thus helping students to experience and assimilate observations into personal conceptions of reality.

It is challenging to teach infopreneurship. When considering the curriculum of a course in infopreneurship, one should look at the knowledge and skills an infopreneur would need in order to be successful. The course content should focus on two aspects, namely, the teaching of creative and analytical skills on the one hand and teaching business basics on the other.

The section on creativity and analytical thinking should include programmes to stimulate ideas generation and the ability to identify opportunities in the market. The section on business basics should include: entrepreneurship theory, definition and profile; how to draw up a business plan; forms of enterprise; legal matters and licensing; basic training in the various functional areas such as marketing, financial and administrative management. The major objective of the course should be to create enterprise capability and, in particular, self-reliance. Enterprise capabilities would involve creativity, initiative taking, persuasiveness, leadership, analytical ability, imagination, networking, moderate risk-taking and achievement motivation.

By implication the end-product should be a creative individual who understands how to establish and operate an information enterprise. It would be unrealistic to expect a relatively short course in infopreneurship to achieve such ambitious objectives. Education for infopreneurship should be a long-term challenge. The contribution of the university is restricted to a rather small section of this educational process.

Presentation methods

The offering of infopreneurship courses requires a radical change in the thinking, culture and environment of Information Science departments at universities [8]. The very nature of the entrepreneur, which is action-orientated, requires that courses on infopreneurship be closely linked to the real world. The role of the infopreneurship lecturer should therefore rather be that of a facilitator, coach and mentor. To achieve a more practical approach to the teaching of infopreneurship, the presentation methods could, among other things, include speeches by infopreneurs, videotapes, creativity, innovation, product development projects and case studies [8]. Research by Holoviak, Ulrich and Cole [9] concluded that entrepreneurially minded students prefer a more active learning style. Significant methods include field projects, role-playing, management simulation, structured experimental exercises and experiments. The possibility of differentiating infopreneurship programmes according to stages of development, career paths and regional and cultural demands could be considered [10].

Research method

Since 1996 the Rand Afrikaans University, situated in Johannesburg, South Africa, has offered a BA (Information Science) degree. Information Science is a compulsory major subject and a further major subject must be chosen from Business Management, Economics, Informatics, Industrial Psychology or Communication Science. This degree has been specifically designed to equip undergraduates with the know-how and skills to develop as infopreneurs. The Department has a business approach in teaching Information Science with the emphasis on information as a resource. The target market is the end-user, that is, any employee or person who needs information for effective decision-making, especially electronic information available on the Internet/World Wide Web. Infopreneurship is a semester course in the second year of study. Students who attend this course can be divided into two groups: students registered for a BA (Information Science) degree with Information Science as a major, and students registered for a BSc (Information Technology) degree with Computer Science as a major. Students are taught how to:

- be creative yet practical in their search for business ideas;
- research their business idea and ensure that it will be profitable;
- develop information products suitable for the market;

- plan the way they will be doing business and handle important issues such as financing the start-up and marketing their products or services.

Students who attended the 1998 course were asked to complete a questionnaire. The survey was conducted towards the end of the course. As this was only an exploratory study, no claims of generalising the findings to other universities are made. Of the sample of 107 students, 83 students (77.6%) completed the questionnaires; 42 were Information Science students and 41 Computer Science students. The responses of the two groups were recorded and averages calculated. A Likert scale was used in ranking opinions.

Findings

Profiles of Information Science and Computer Science students

The average age of both groups was 19 years. Forty-five per cent of the Information Science students were male compared to 67% of the Computer Science students. Both groups were more positive about starting an own information business after completion of the infopreneurship course, with 61% of Information Science students and 83% of Computer Science students stating that they were more positive after the course.

Objectives of a course in infopreneurship

The respondents had to indicate on a Likert scale, from 1 = irrelevant to 5 = extremely important, the importance of a given set of objectives. In order of preference, the average rankings were as illustrated in Table 1.

The most important difference occurred with regard to assisting students to become self-employed, which was placed first by Computer Science students and fourth by Information Science students. This perception of Computer Science students may stem from the strong competition in the computer industry in South Africa. Computer Science students regarded the contribution of the course towards an entrepreneurial culture as

Table 1. Objectives of a course in infopreneurship

Objectives of course	**Average ranking: Information Science students**	**Objectives of course**	**Average ranking: Computer Science students**
To create entrepreneurial awareness	3.9	To assist students to become self-employed	4.2
To teach the know-how to start and run a business	3.9	To create entrepreneurial awareness	4.0
To prepare students to start their own businesses	3.7	To teach the know-how to start and run a business	4.0
To assist students to become self-employed	3.7	To prepare students to start their own businesses	4.0
To develop analytical and creative skills	3.7	To contribute towards an entrepreneurial culture	3.9
To develop students into intrapreneurial employees	3.6	To develop analytical and creative skills	3.9
To contribute towards an entrepreneurial culture	3.5	To develop students into intrapreneurial employees	3.5

less important than Information Science students did. Comments by Computer Science students stated that they already see themselves as entrepreneurs who are going to start their own businesses.

Course content

In this section respondents had to indicate the importance of the inclusion of 18 different topics in an infopreneurship course. The Likert scale used varied from 1 = irrelevant to 5 = extremely important. In order of preference, which was based on the average rating, course content was rated as shown in Table 2.

The most important differences between the two groups occurred with regard to the improvement of creative ability and improvement of analytical thinking. In both instances Information Science students expressed a greater need for training in these areas than Computer Science students did. The perception of Computer Science students that they had a lesser need for training in creative and analytical skills may stem from the emphasis in the Computer Science department on creativity, analysis and design. It appeared as though Information Science students did not regard themselves as very creative.

Both groups gave a low ranking to definition and theory of entrepreneurship. This was in accordance with the finding of the literature survey; namely that entrepreneurship should be practical. Examples of infopreneurship received a much higher ranking from Computer Science students than from Information Science students.

Table 2. Course content

Content	Average ranking: Information Science students	Content	Average ranking: Computer Science students
Financial management	4.1	Risk and the entrepreneur	4.3
Improvement of analytical thinking	4.0	Infopreneurship in the new South Africa	4.0
Improvement of creative ability	4.0	Marketing management	4.0
How to draw up a business plan	3.9	Examples of infopreneurship	4.0
Ideas generation techniques	3.8	Profile of the entrepreneur	3.9
Risk and the entrepreneur	3.7	How to draw up a business plan	3.9
Financing the new venture	3.6	Improvement of creative ability	3.9
Infopreneurship in the new South Africa	3.5	Evaluation of new ideas	3.9
Profile of the entrepreneur	3.5	Ideas generation techniques	3.9
Market research	3.5	Financial management	3.8
Evaluation of new ideas	3.5	Improvement of analytical thinking	3.8
Examples of infopreneurship	3.5	Sources of new ideas	3.8
Feasibility analysis of new ideas	3.4	Personnel management	3.7
Definition and theory of entrepreneurship	3.3	Financing the new venture	3.7
Sources for new ideas	3.2	Market research	3.7
Marketing management	3.1	Feasibility analysis of new ideas	3.7
Personnel management	3.0	Intrapreneurship	3.5
Intrapreneurship	3.0	Definition and theory of entrepreneurship	3.4

Table 3. Teaching methods

Teaching method	**Average ranking: Information Science students**	**Teaching method**	**Average ranking: Computer Science students**
Lectures on infopreneurship	3.5	Group visits to information businesses	4.0
Practical assignments	3.5	Lecturer relating own infopreneurship experience	3.9
Research paper assignments	3.5	Business plan preparation assignment	3.9
Lecturer relating own infopreneurship experience	3.5	Practical assignments	3.8
Business plan preparation assignment	3.5	Lectures on infopreneurship	3.7
Case studies	3.5	Speeches by infopreneurs	3.7
Creativity, innovation and product develop exercises	3.4	Case studies	3.7
Group visits to information businesses	3.4	Research paper assignments	3.5
Text reading and discussions	3.2	Text reading and discussions	3.4
Speeches by infopreneurs	3.1	Creativity, innovation and product development exercises	3.3
Videotapes	2.8	Videotapes	3.2

The Information Science and Management Science background of the lecturer, and consequently the more frequent use of examples of entrepreneurs with Information Science training, could be the reason. The most important differences between the two groups occurred with regard to the teaching of financial management and marketing management. Financial management was placed first by Information Science students and tenth by Computer Science students, and marketing management third by Computer Science students and sixteenth by Information Science students. No logical explanation can be provided since the majority of students of both groups also attended Business Management courses. Both groups gave a low ranking to the teaching of intrapreneurship, which is an indication that students already decided that they would rather work for themselves than in an existing enterprise.

Teaching methods

As already mentioned, it appears to be difficult to teach infopreneurship effectively. In this section of the questionnaire respondents were requested to indicate the most effective methods to stimulate entrepreneurship abilities among students. The teaching methods were ranked on a Likert scale, where 1 = totally ineffective and 5 = extremely effective. The average ranking is given in Table 3.

In similar fashion to the response with regard to course content, creativity, innovation and product development were rated much higher by Information Science students than by Computer Science students. The higher perceived need for such teaching among Information Science students most probably stemmed from less training in the field of creativity and innovation than amongst Computer Science students.

The lecturer relating own infopreneurship experience, and practical and business plan assignments were ranked high by both groups. This confirms the requirement stated in the literature that entrepreneurship courses should be practically orientated and that the lecturer should have practical experience in applying infopreneurship.

Both groups expressed strong dislike for the use of videotapes as a teaching method. The very poor ranking of videotapes was most probably caused by the fact that there is still much room for improvement in the quality and appropriateness of the videotapes available for use in infopreneurship courses. This area surely offers a business opportunity for a prospective entrepreneur!

Conclusion

The introduction of infopreneurship at universities presents a number of serious problems, such as finding suitable staff and finding an appropriate mixture of often demanding and expensive teaching methods, for example business games, videos, case studies, guest speakers and establishing interaction with enterprising business people.

Infopreneurship courses should be adapted to the unique needs of the target group they are offered to. As stated, the courses should in the first instance be practically orientated. More innovative methods need to be developed to make the presentation of infopreneurship more interesting, and possibly more practical, and there still appears to be a need for high quality and relevant videotapes on infopreneurship.

An education system that places a premium on knowledge acquisition and retention – while it restricts creativity, capability and the ability to relate to others and conditions young people to be dependent on others for employment opportunities – is out of phase with the needs of society in the new millennium. Although researchers argue that as few as 15% to 20% of the individuals in any population have moderate to strong entrepreneurial abilities, students who attend infopreneurship courses will definitely have a fuller understanding of the entrepreneurial phenomenon, and become more competent stakeholders – whether as employees in the information industry or, possibly, infopreneurs.

References

1. Koenderman, T. *Not just a bunch of ne'er-do-wells*. 1998. http://www.fm.co.za/98/0904/focus/young.htm

2. Peters, R.F. Information partnerships: marketing opportunities for information professionals. *Information Outlook*, 1(3), 1997, 14-16.

3. Weitzen, H.S. *Infopreneurs: turning data into dollars*. New York: John Wiley and Sons, 1998.

4. MaCleod, G. *Starting your own business in South Africa*. Cape Town: Oxford, 1995.

5. Zeihaml, C.P. and Rice, G.H. Entrepreneurship/small business education in American universities. *Journal of Small Business Management*, 25, 1987, 44-50.

6. Kotze, W. and Boon, J.A. Library and information brokering and consultancy in South Africa – an exploratory study. *South African Journal of Library and Information Science*, 63(3), 1995, 114-121.

7. Dana, L.P. An international survey of entrepreneurship education. *Journal of Enterprising Culture*, 1(1), 1993, 67-92.

8. Benson, G.L. A note on entrepreneurship education in colleges and universities. *South African Journal of Entrepreneurship and Small Business*, 1(3), 1989, 43-52.

9. Holloviak, S.J., Ulrich, T.A. and Cole, G.S. Training entrepreneurs. *Performance and Instruction*, Nov-Dec, 1990, 27-31.

10.Ronstadt, R. The educated entrepreneurs: a new era of entrepreneurial education is beginning. *American Journal of Small Business*, 11, 1987, 37-53.

CONTENTS

93–97 Comment
Why isn't the Internet secure yet?
J.D. Tygar and Alma Whitten

98–114 The impact of the Internet on information seeking in the Media
David Nicholas, Peter Willliams, Peter Cole and Helen Martin

115–123 Citizenship information needs in the UK: results of a national survey of the general public by personal doorstep interview
Rita Marcella and Graeme Baxter

124–129 Information retrieval techniques for evaluating search engines: a critical overview
Monica Landoni and Steven Bell

Comment
Why isn't the Internet secure yet?

J.D. Tygar
University of California, Berkeley – SIMS, 102 South Hall 4600, Berkeley, CA 94720–4600, USA
tygar@cs.berkeley.edu

Alma Whitten
Computer Science Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213–3891, USA
www.cs.cmu.edu/~alma

Futurists promise us that the Internet will soon be the premiere vehicle for communication, information management, commerce, entertainment, and education. Citizens of the world will be united in a single, easy-to-use, universal network that will deliver a new renaissance of democracy and free flow of knowledge, fulfilling the promise of a global village. While these predictions reflect a certain degree of technological optimism, even the most jaded observer will concede that the Internet has fundamentally changed – and will continue to change – the way we access information. And the full flowering of this new golden age will occur just as soon as we make the Internet safe and secure.

But researchers have been working on network security for at least 30 years – and they have delivered powerful technologies: new technologies for cryptography protecting information that is transmitted or stored in a distributed system; new technologies for access control that allow information owners to exactly specify who can access that information; new technologies for electronic commerce that allow intellectual property owners to charge for access to information; and new technologies for protected execution environments that allow remote code to run safely on a host system. And while fundamental research problems still exist, we have technology that is perfectly adequate for securing most contemporary applications.

So why isn't the Internet secure yet?

We argue that a fundamental problem is the 'usability of security'. Even when applications have powerful mechanisms built in they are so difficult to use that most users do not properly configure their security systems. Vendors attempt to make security easy to use by causing pop-up boxes to show up in dangerous or anomalous situations – but the response of most users is to take any action necessary to make the annoying pop-up box disappear. Indeed, one clever programmer has even written software to automatically click 'OK' on all pop-up boxes so users of the software will be untroubled by system notices.

But security is not simply a function that can be automated away – it depends heavily on the context and content of the information that is to be protected. While human beings can understand the difference in the secure needs associated with files containing medical records as opposed to files containing athletic records, it is not clear that machines will understand this. Moreover, as new threats emerge, machine protections will always be out of date. Worse of all, by automating security, one provides attackers with an algorithm for understanding exactly what steps will be taken in case a system comes under attack. So we need to find a way to make users able to be responsible for their own security.

Perhaps the greatest single weapon in the arsenal of security tools is 'cryptography'. As is well reported in the literature, cryptography has the power to provide secure communications, protect transactions, provide powerful privacy, and validate the integrity of information. Cryptography forms a formidable tool, but one that will lie fallow if people do not know how to use it. (We have already heard numerous reports that most users are not able to understand – or do not regularly inspect – the underlying cryptographic certificate mechanism offered to show the authenticity of web pages and applications.)

Understanding security is made more difficult by the complex melange of different mechanisms used by a variety of applications and web sites. Consider the problem of integrity of software. Perhaps you are careful to only install applications originating from trusted sources. You may be surprised to learn that your system is still vulnerable to attack. Powerful applications such as word processors, spreadsheets, and presentation software not only display content but run active programs. These active programs introduce a set of potential risks – it is all too easy to include a computer virus in a Microsoft Word file, for example. If the problem is serious, patches and fixes are inevitably provided by vendors, but require diligence to install them on a regular basis. And these are just stopgap solutions – to safely run executables loaded over the network, you need to be able to supervise and control the behavior of programs on your system. One solution is for you to never view any content that you did not create, but to take such a stand is to abandon the vision outlined at the start of this article.

Or, consider the problem of privacy. Most readers of this article have probably ordered a book from an online bookseller such as Amazon.com. Were your purchases of those books protected? How can you find out?[1] The answers will be different for each web merchant you transact business with. How can you tell what a site is really doing as opposed to its public stance? In most cases you cannot. (These problems have become quite pervasive – the following story is a powerful anecdote of the threats inherent in such loss of privacy: a woman filed a personal injury lawsuit against a major California supermarket chain after she slipped in a puddle of water and had to be hospitalized with a broken hip. The supermarket is alleged to have threatened to disclose in court the customer's purchases, as recorded from her supermarket frequent buyer card, with the intent of showing that she had purchased a large amount of alcohol and was thus likely to have been drunk at the time of the accident.) The solution to these privacy problems is particularly challenging since it requires a combination of clever technology together with a clear, coherent policy stance.

The Internet continues to grow at a frightening rate. At this stage of the Internet's development, most new users to the Internet are not very technologically sophisticated. Thus, the average level of awareness of security and security mechanisms is very likely decreasing unless we take active steps to make users more security savvy across the Internet.

Of course, problems of hard-to-understand software and poor user interfaces are ubiquitous across the field of computing. So why are we putting so much attention on the usability of security and privacy mechanisms? Because the stakes are so much higher with security:

- Security software requires that users understand all mechanisms – even a single weak link in the chain can lead to compromise of information. This is quite different from most software, which does not require a user to understand all mechanisms. For example, perhaps the reader has never needed to explore the intricacies of mathematical typesetting or the use of Japanese characters in his or her favorite word processor. You will not need to understand these until you have to use them. But if you do not completely understand how to generate, protect, and use your cryptographic keys, your information will be subject to attack. For firms, a single employee who does not understand the mechanisms can be a risk to the firm's entire site.
- Once disclosed, private information can not be recovered and made private again.

---

[1] As of October 1999, the answer for Amazon.com can be found after some searching of their web site. The company reserves the right in the future to freely distribute purchase information to third parties unless you send blank email messages to never@amazon.com from each account you use to shop with Amazon.com.

One can protect a site after it has been attacked, but this is rather like closing the barn door after the cows have left. This is different from most software, which is more forgiving – for example a mistake made in editing a document on a word processor can always be fixed at a later date.

- Users may not even realize that something is amiss with the security of their system until a disaster strikes. For example, if your communications are subject to eavesdropping, you probably will not realize it until those communications are monitored and then disclosed. Even more frightening is the prospect that communications are monitored and not disclosed. This is different from most applications, which more immediately reveal when a problem exists. To again use the example of a word processor, a document typesetting error will usually be quite obvious to the user.

How bad are things, really?

Nobody claims that computer security is easy for people to understand and manage, but how terrible are these problems in real, day-to-day life? When an ordinary person tries to protect her privacy and security while using the Internet, does she face a task that is tediously confusing and inconvenient, but one that can be managed if necessary? Or is it the sad truth that most security software is simply impossible for most people to manage at all?

To get some answers to these questions, we assembled a group of twelve people who were experienced users of computers and particularly of email. All twelve had at least some college, and several had advanced degrees. Some were artists, some worked in medicine, some were administrators, and some did computer programming. Our reasoning was that if this group of diverse, yet relatively competent and accomplished, people could not successfully make their email secure and private, then the outlook was bleak for the general population to be able to do so.

We gave these people the software that had the most elegant graphical user interface for email security and privacy: PGP 5.0 combined with the well-known email program Eudora. We gave them a printed user manual for them to consult as necessary. Then we gave them a secret email message and a list of email addresses, and asked them to send the secret message, digitally signed and encrypted, to the addresses on their list.

The results were enlightening, to say the least[2]. Three of them promptly sent the secret message off without any encryption to protect it, mistakenly believing either that they had encrypted, or that the software would magically and invisibly apply the encryption on their behalf. One did not even realize his error after the fact.

One person never managed to figure out how to encrypt at all, even after an hour and a half of trying. Seven of the others became confused about the basics of how public key cryptography works, and used the wrong keys to encrypt, so that the people they sent the secret message to could not decrypt it. Only half of the group managed to eventually send any correctly signed and encrypted email messages, and only four of them succeeded in sending the original secret message, digitally signed and encrypted, to the addresses on the list as we had asked. Many problems, pitfalls, and points of confusion had to be overcome en route to this limited success, often leading to the expression of great frustration.

We wanted to know how bad the problem was: whether generally competent people could manage email security if they were willing to put in reasonable effort, or if even the best-designed existing software was unrealistically hard to understand and use. The answer seems clear: the problem is bad. Security is so unintuitive and confusing to most people that we need a new vision of how to make it usable: something beyond the kind of user interfaces that work acceptably for word processors and spreadsheets. Something very simple and clear, something straightforward and definite in its depiction of how each security mechanism works and how to make use of it.

---

[2] For those who are curious, a formal report of this research is available in *Why Johnny can't encrypt: a usability evaluation of PGP 5.0* [1].

What the future holds

As problems in network security continue to proliferate, we believe that pressure will grow for better security user interfaces. But, given the challenges outlined above, are better user interfaces really possible? We believe so. Here is the set of design strategies we think are most promising:

- ***Simplify, simplify, simplify***
  Security software is not immune to the market-driven process of feature creep, which causes programs to increase in size and complexity with each successive release, as the latest variations and buzzwords are incorporated into the functionality of the program. Nor is it exempt from the desirability of catering to the needs of a wide range of users: from the simple needs of novices to the complicated and advanced requirements of experts. Size and complexity, however, work against the crucial goal of designing security that the maximum number of users will be able to understand quickly, clearly and effectively.

  How can security software be made simple to understand while still allowing for the needs of a wide variety of users (and of marketing departments)? By carefully identifying just what practical use beginners will actually need to make of the security functions we can construct a basic and coherent conceptual model of the security mechanisms that they will need. All other functionality and data can be pushed into one or more levels of background, so that the advanced features are available to experts but do not confuse or overwhelm beginners. Structuring the functionality of the security into multiple levels that correspond to clearly identified conceptual models also allows us to make the best use of our other design strategies, as we describe below.
- ***Communicate clearly***
  Once we have the security functionality structured into multiple levels, we are well placed to focus on how best to quickly and effectively communicate the workings of the novice level to users. The security mechanisms in the novice level must be presented in terms of a unifying and accessible conceptual model, which should be reinforced as each aspect of the security system is displayed.

  Visual metaphors are of obvious use in presenting such a model, but need to be carefully constructed and tailored so as not to mislead users. In PGP, it is essential to know how certain crucial information used to configure cryptography (the cryptographic keys) are used. And, of course, it is essential to understand the operations done on data to protect it. But PGP does not help very much with understanding the keys or the operations. The icons used to represent the cryptographic keys and the encryption operation did not help the users understand public key encryption well enough to know which keys to use, and the quill pen icon used to represent digital signatures and the signing operation did nothing to tell the users that keys are used for signing. Users would have been much better served if the designers had focused on conveying the basic conceptual model of public key cryptography and tailored their metaphors to reference and reinforce that model.

  After designing an initial user interface that optimizes the communication of the necessary basic model, then we can attempt to design the remaining levels to gracefully expand the model as the user comes to need more complex security functionality.
- ***Guide and protect***
  A concept from education and, recently, from research in educational software, is 'scaffolding': the technique of providing pervasive and explicit help in the early stages of a learning process, which is then gradually withdrawn (a process referred to as 'fading') as the learner acquires skill and understanding. Our structuring of security functionality into multiple levels which are only gradually accessed by users is itself a form of scaffolding, in which the beginner is helped by being given only a limited and simplified model, to protect against the

risk of confusion and error. Scaffolding of this type is sometimes called 'training wheels'.

Other types of scaffolding are also necessary and appropriate for security user interfaces and the multi-level structuring we propose provides a natural framework on which to build and fade additional guidance and help. Warning messages and other kinds of brief tutoring (in the form of interruptions) that can be presented aggressively to novices and faded as users become more expert. Users can be allowed to turn off such messages as they become annoying, providing another natural fading mechanism, if the messages first provide educational content and are not merely risk flags. This last point is central and is sadly lacking from much contemporary software.

Will the Internet be secure?

The march towards moving mass information systems to computer networks seems unstoppable. Every day, we read in the paper (or even better, on our favorite electronic news site) about further applications and information repositories available on the World Wide Web. But in ten years, will the Internet be significantly more secure than it is today? We hesitate to answer, since speculating on the future development of the Internet at times resembles astrological forecasting. But of one thing we are certain, we will be paying a cost. Either we will pay the cost to build and deploy more effective, usable security systems, or we will pay the cost in the form of lost privacy and security. It is easier for an individual to protect privacy from the beginning than to try to recover it once it has been lost. The same holds true for a society.

References

1. *Why Johnny can't encrypt: a usability evaluation of PGP 5.0.* Proceedings of the 8th USENIX Security Symposium, August 1999. Additional information is also currently available at: http://www.cs.cmu.edu/~alma.

The impact of the Internet on information seeking in the Media[1]

David Nicholas, Peter Williams, Peter Cole and Helen Martin
Internet Studies Research Group, Department of Information Science, City University, Northampton Square, London EC1V OHB
nicky@soi.city.ac.uk

There is very little qualitative data on what impact the Internet is having on information seeking in the workplace. Using open-ended interviews, questionnaires and observation, the impact of the Internet on the British Media was assessed. The focus was largely on newspapers, with *The Guardian* being covered in some depth. Over 300 journalists and media librarians were surveyed. It was found that amongst traditional journalists use was light. Poor access to the Internet – and good access to other information resources – were largely the reasons for this. Of the journalists it was mainly the older and more senior journalists and the New Media journalists who used the Internet. Librarians were also significant users.

Searching the World Wide Web was the principal Internet activity and use was generally conservative in character. Newspapers and official sites were favoured, and searches were mainly of a fact-checking nature. Email was used on a very limited scale and was not regarded as a serious journalistic tool. Non-users were partly put off by the Internet's potential for overloading them with information and its reputation for producing information of suspect quality. Users generally dismissed these concerns, dealing with potential overload and quality problems largely by using authoritative sites and exploiting the lower quality data where it was needed. Where the Internet has been used it has not been at the expense of other information sources or communication channels, but online hosts seem to be at most risk in the future.

Introduction

The Internet has become the information world's obsession. Plainly it poses many strategic and searching questions for information users, information professionals and the information industry alike. However, we have little detailed knowledge on how it is impacting on information seeking in the workplace. The British Library funded research project – The Changing Information Environment[2] – set out to gather such data. The chosen ground for the investigation was the Media. The Media was chosen because:

- With information so central to its activities it might be expected that it would be amongst the 'early leaders' in the adoption and adaptation of the new technology.
- The pressurised and demanding nature of newsrooms offered an ideal environment in which to investigate the impact of the Internet unobtrusively. In the Media, events are telescoped, resources are relatively plentiful and impacts (if any) are relatively immediate, thus providing ideal laboratory conditions for impact research.
- The Media was the subject of a previous British Library Research and Innovation Centre study[3] that sought to provide an early insight as to how end-user access to full-text, natural language databases would

---

[1] Based on the Report of the British Library funded research project The Changing Information Environment: the Impact of the Internet on Information Seeking Behaviour in the Media. Published as *The Media and the Internet* [1].

[2] April 1997–April 1998.

[3] Information seeking in an information society: the end-user. 1985–1987.

impact on the information profession. That research concluded – correctly, as it turned out – that end-use would not prove the information turning point. The impact on the information seeking behaviour of journalists was light and the threat to the intermediary proved to be largely bogus [2]. Possibly it would be different this time around – a decade later, and with a much more powerful information system in the reckoning?

Within the Media it was largely the press that was investigated: tabloid and broadsheet national newspapers, English and Scottish and regional newspapers were covered, as well as some news-based magazines. One paper, *The Guardian*, was investigated in some depth to obtain a detailed insight into the situation at one site. A number of broadcasting organisations were covered, most notably the BBC. Table 1 provides a list of the names of the major organisations surveyed.

The main aim of the research project was to examine closely and qualitatively the impact of the Internet on a strategic information community to whom it might be assumed to be of immense significance. The key research questions were these:

1. How – and by whom – was the Internet being used in the Media?
2. What were the key information seeking considerations associated with its use – did for instance quality and quantity issues weigh heavily on the users?
3. What impact, if any, was the Internet having on the use of traditional existing information services and communication channels, and especially libraries?

Related research

What we know about the impact the Internet is having on the Media comes largely from the US. The topic is widely discussed but little researched in the UK. Most of the US studies feature the ubiquitous Internet-delivered questionnaire surveys. Apart from their inherent bias towards users of the Internet these surveys are limited in that they tend to be concerned mainly with quantifying the increase in Internet use in the Media [3]. What these studies do point to is the Internet's growing influence on the US Media. One study, for instance, claimed that 89% of those journalists with access to the Internet have used it, and 45% made use of the Internet on a daily basis [4]. However, there are plainly some distinctly national factors at work:

1. Internet use in the US is stimulated by free local telephone calls.
2. Local and Federal Government data is freely and widely available on the Web.
3. In the US there are large numbers of 'beat' reporters employed by large local newspapers – all with a narrow, local brief, and, what with the aforementioned availability of local information, a strong need to search the Internet.
4. Print journalists in the US are not slaves to a particular style of newspaper reporting, as they often are in the UK, and, in consequence, can incorporate the Internet more easily into their reporting.

Computer-aided journalism (CAR) – something that has led to dramatic changes in news gathering in the US according to Paul [5] – appears to have failed to take root in the UK. CAR can only thrive in an environment where information is liberally and generously placed in the electronic public domain. This is a long way from happening in the UK.

However, even in the United States the Internet has some way to go before it displaces traditional journalistic methods. It is largely used as an adjunct to existing systems of communicating and gathering information. Thus, one of the biggest surveys of US journalists [4], found that none of the Internet sources – listservs, email, the World Wide Web and Newsgroups – were cited by more than 9% of journalists as their primary source of stories. Also disciples of Computer-aided reporting, like Garrison [3], have admitted that even in the favourable climate of the US CAR has a marginal impact on newspaper reports. Possibly, the most interesting finding to come from the US also comes from Garrison. Apparently online in general and the Internet in particular are leading to growing end-use and diminishing intermediary use. However, the explanation for this could lie, not so much with the Internet, as with the growing use in the US Media of personal researchers, working alongside journalists – something that is yet to happen in the UK.

Research design

Depth, open-ended interviews were the preferred research tool because they lent themselves so well to the investigation of the new, the changing, the controversial, and the ill-defined [6]: the topic being investigated was all of these things. Inevitably, too, not just the issue being studied, but also who was being investigated determined the choice of methods. Thus, while there was little chance of persuading very busy journalists to subject themselves to the kind of examination that information researchers subject academics to – e.g. diaries, questionnaires and talk-aloud techniques at the terminal [7] – there was every chance of persuading them to be interviewed. 176 journalists and media librarians were interviewed in this way. Closed question interviews and group interviews were used also to obtain more quantitative data (adding another 74 people to the interview sample). Questionnaires (79 returned) and observation (16 people observed) were also employed but on a limited scale and largely to buttress and contextualise the interview data. Altogether data were gathered from around 350 people from more than 50 media organisations in the UK (Table 1).

Obtaining a scientific and systematic

Table 1. Methods employed and size and origin of sample

Organisation	**Journalists**			**Librarians**			**Others**[1]
	Interview	Questionnaire	Observation	Interview	Questionnaire	Observation	Interview
BBC	7			17	4		2
Eastern Counties Network	1			1			1
The Guardian	36	2	7	14	1	5	
Herald (Glasgow)	6		2	5	5		
London Weekend Television	0			2			1
Mirror	4			1			
News International	50[2]			15	11		1
Newsquest Northeast	3			3	2		1
The Observer	7		1				
Scottish Daily Record	1	1		5			1
Telegraph	4			1	1		
Time-Life	4			1			
University of Central Lancashire	1	33[3]					
VirginNet	4			1			
Freelancers	6						
Various local newspapers[4]	26				7		
Other media organisations	12		1	4	12		4
Total	**171**	**36**	**11**	**68**	**43**	**5**	**11**

1. Systems/IT editors; 2. Internet question as part of general library interviews; 3. Journalism students; 4. Includes group discussion with 24 trainee journalists.

sample was not possible given the size, heterogeneity and high-pressure lives of the population investigated. However, there was a need to ensure that the various sectors of the Media industry, the various types of journalist, end-users and librarians, and users and non-users of the Internet, were represented. The Association of UK media librarians' (AUKML) membership directory was used to contact librarians, and with the help of these librarians, a sample of journalists was obtained. It really was very much a case of those that were interviewed were those willing to be interviewed – and Internet users and enthusiasts tended to be the most obliging. The danger of obtaining a self-selecting, biased sample was recognised and questionnaires and observation were employed to compensate for this. Also a detailed case study was undertaken at *The Guardian* (a quality daily paper).

Most of the interviews were of an open-ended question type and generally lasted about 45 minutes. The technique had been developed by the senior author during user needs interviews with journalists over a number of years [6]. The intention was that detailed information about the Internet should emerge from general questioning about their work and the impact (or potential) of the Internet on it. However, there were topics we wished to explore and information that we needed to gather and if this did not come naturally users would be gently and slowly prodded towards covering/providing it. The broad areas of interest were:

Background: Gender; job role; department; age, computer competency/experience.
Use/non-use: Volume, frequency etc.
Nature/purpose of use: What it is used for – communicating, fact-finding, browsing etc?
Possible reasons for non-use: Unavailability, response time etc.
Satisfaction/value of Internet data.
Sources/sites visited: examples of searches.
Searching methods.
Impact on/relationship with existing sources/systems.
Overload.
Impact on work.
Training.
Intermediary/library: Do they delegate the search?

Other forms of interview used

Two other forms of interview were employed. Both were intended to increase the journalist sample. Two discussion/focus groups were held with 24 journalists from local newspapers in the North of England that were attending a training course. They covered the same ground as the depth interviews. The other type of interview was a collaborative effort between the research team and librarians at News International. This was a closed question interview that was held with 50 *Times* journalists. Journalists were interviewed about the Internet as part of a general library survey. This was a 20% sample and all those asked agreed to be interviewed. The intention was to obtain some quantitative data about Internet use in one particular location – something the project lacked at the time.

Observation

The observation of journalists was an integral aspect of the fieldwork and was used in two ways. In the first instance it was used to prepare the ground for the interviews. There was a need to witness the whole information process in motion and place Internet use in context. Thus observation initially provided an understanding of the circumstances surrounding the processes and issues being studied, and helped establish the credibility of the investigator. In the second instance observation was used as an adjunct to the interview to increase and improve the data yield. Through observation it was also possible to establish general levels of Internet use and establish where the Internet was being used. Most of the observation took part at *The Guardian.*

Questionnaires

Questionnaires played a minor, but important role in the study. They were largely used to obtain data from librarians about both their use of the Internet and their organisation's use of the Internet. The following questions were asked. Do you have access to the Internet at work? Do you use the Internet at work? How much use do you make of the Internet at work? What kind of use do you make of the Internet? Do journalists at your organisation have any access to the Internet at work, apart from through the library? How many journalists have access on their own desk top? How many journalists have access within their department or close by? The questionnaire was distributed to libraries on the AUKML mailing list and a few notable absentees from this list – 43 (50%) were returned. A questionnaire was also used to discover whether student journalists were using the Internet. The questionnaire was distributed to a class of student journalists at the University of Central Lancashire – 33 (90%) were completed.

Results

Use of the Internet

The Media was chosen for study because it was felt that journalists, as information seekers and packagers *par excellence*, might be expected to be in the advanced guard of Internet users. All the indications from the United States was that this was the case [4]. As this was essentially a qualitative study, no concerted attempt was made to gather national use data. However, piecing together the quantitative use data we did obtain, use appeared to be generally low and patchy in the UK. Thus:

- A closed question interview survey of 50 journalists from *The Times* (a quality daily paper) showed that over two-thirds did not use the Internet (Table 2).
- A group interview with 24 trainee journalist from various local and regional newspapers showed that use was low – just four (16%) used the Internet.
- Although 68% of the 52 people interviewed during *The Guardian* case study were Internet users, this represented less than a quarter of all journalists at the paper, and there was a strong bias towards Internet users amongst our sample. Comments from those we interviewed and the results of our own observation bore this out. This is what the Editor had to say about overall levels of use at his newspaper. 'I had lunch with the entire staff last year – twenty times – and at the end of every session I got them all to put their hands up and say whether they were using it or not and you tended to get about four out of twenty [20%].'

Table 2. Use of the Internet by Times journalists

Frequency of use	*n*	%
Several times a day	2	4
Once a day	4	8
Several times a week	4	8
Once a week	3	6
Once a fortnight	2	4
Variable	3	6
Didn't use	32	64
Total	**50**	**100**

There were three groups who used the Internet quite extensively: student journalists, media librarians and New Media journalists – those journalists working on newspaper web sites. Benefiting, no doubt, from academe's generous Internet provision, student journalists searched the Internet with some alacrity, with 97% (32) of them saying they used it and 58% (19) claiming to do so very frequently (Table 3). While a good proportion of journalists still have to make up their mind – or have their minds made up for them – regarding the journalistic value of the Internet, their librarian colleagues had no such doubts: 94% (47) of those interviewed and 74% of those returning questionnaires used the Internet (Table 4). Of those surveyed one-third used 'it all the time'; 28% 'used it heavily some of the times' and the rest (39%) 'used it occasionally'. Roughly the same

Table 3. Use of the Internet by student journalists

At college	*n*	%
Very frequently	19	58
Sometimes	8	24
Occasionally	5	15
Never	1	3
Total	**33**	**100**

Table 4. Use of the Internet: depth interview data

Interviewees	National and regional newspapers		Specialist publications and freelancers		New Media		All journalists		Librarians		Total	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Users	34	60	9	82	13	93	56	68	47	94	103	78
Non-users	23	40	2	18	1	7	26	32	3	6	29	22
Sample	57	100	11	100	0	0	82	100	50	100	132	100

proportion (93%) of the New Media journalists were Internet users.

Factors influencing use

Access

Lack of Internet access was a major factor in the low and uneven Internet take-up – nearly a third of journalists surveyed had no access at all to the Internet. The poor provision, certainly in the case of the national newspapers, was largely due to the fact that the editorial systems on which journalists conduct their work – and, in some cases, access to commercial online systems like FT Profile – were not Internet terminals. Even where there was access, via stand-alone PCs, it was generally remote – with journalists having to walk to the far end of a large office, or even go to another floor, to find a terminal. The situation was made worse sometimes by the lack of a printer. Over half of all non-users indicated that they would use the Internet if access were improved. Those with easy access – senior reporters on national newspapers, those working for some specialist publications (who tended to work in networked PC offices), news librarians and freelancers, who had invested in home Internet terminals – tended to be the heavier users.

Age experience

If the relatively poor Internet take up in the workplace was unexpected, another surprise was the chief characteristics of the users. Far from being the stereotypical young, most (66%) were experienced journalists into their thirties/forties (Table 5). Young, IT literate post PC graduates were supposed to flood the workplace, expecting to use the Internet as a matter of course. However, young journalists were not even the strong proponents expected, although the survey of journalism students

Table 5. Internet use by age[1]

Age	National and regional newspapers		Specialist publications and freelancers		New Media		All journalists		Librarians		Total	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
up to 29	6	18	0	–	5	38	11	21	15	33	26	26
30–39	5	15	5	71	5	38	15	28	19	41	34	34
40–49	16	48	2	29	2	16	20	38	12	26	32	32
50–59	6	18	0	–	1	8	7	13	–	2	7	7
60 and over	–	–	–	–	–		–	–	–	–	–	–
Total	**33**	**100**	**7**	**100**	**13**	**100**	**53**	**100**	**6**	**100**	**99**	**100**
Unknown	1		2				3		1		4	

[1] Those who had personally used the Internet on more than one occasion.

indicated that this might change in the future. The age profile of users was explained partly by the age profile of journalists in the workplace, although there are two other explanations. Firstly, it was the 'older hands' that saw the introduction of online ten to twelve years ago, assimilated it into their journalistic practices and were now viewing the Internet in much the same light. Secondly, age and seniority go hand in hand. Seniority bestowed upon the reporter the luxury of a more flexible job role, more autonomy and more job security. These were found to be ideal conditions to germinate interest in the Internet. Contrast this with the situation of the young news reporters who are confronted with a large degree of job uncertainty and anxiety, employed as many of them are on short-term contracts.

Education and training

Logic would suggest that training would be a major determinant in Internet use but this proved to be only partially true. Non-users, who were not ideologically or otherwise against the Internet, invariably lamented the fact they did not have time to attend training sessions or that they had no time to learn. This also appears to have inhibited use among older non-user journalists. Many of those who reported that they happily used online services (from their own desktops) felt that they needed Internet training. As one journalist pointed out: 'I think it is sold to people on the idea that it makes things simple but I think that is as much a 'con' as telling somebody they can learn a language in seven days'. Heavy users, however, were not necessarily those who had been trained. Rarely was training mentioned, and almost never as a formal ongoing process. Many users were entirely self-taught.

Types of use

There are as many uses for the Internet as there are Internet sites or Internet users. Amongst a seemingly endless list of activities, journalists variously claimed to have indulged in a real-time chat with Czech dissident students, accessed OECD statistics, bought government reports, checked the latest information about computer viruses, downloaded free software, kept abreast of current events, and monitored the activities of a religious cult.

Web resources used

The sites favoured by journalists tended to be those offering familiar, traditional and authoritative information: online newspapers and news services; and other 'official sites' – government and other institutions such as universities (Table 6). By far the most popular locations were online newspapers, mentioned by two-thirds of journalists. Such sites were not so popular with librarians possibly because they had better access to alternative online services, such as FT Profile. Librarians tended to prefer unofficial sites, especially entertainment ones.

The extensive use of the Internet to consult other news publications is a natural extension of the journalists' predilection for researching cuttings or newspaper archives such as those of FT Profile. Newspapers of every description were mentioned with the foreign press appearing to be particularly popular. Of those journalists who named specific newspapers or types of papers, over 50 named US titles or the US press generally. *The Washington Post* was the paper mentioned most frequently. Journalists used the newspaper sites for all kinds of purpose: researching for specific information, filling in the history or background to a current issue, keeping abreast of the news in other countries or searching for stories the wires may have missed.

Journalists were far more likely to be verifying facts or researching for specific information on an individual or organisation, than browsing or 'surfing' the Net for any story they could find. The idea that journalists spent their time surfing the Internet for this purpose was laughed at by more than one journalist. Nevertheless, this is what many journalists were doing when they scanned the US or other foreign press. Freelancers and New Media journalists did, however, openly admit to surfing. For the former the Internet was regarded as something of a treasure chest of potential stories, leads, ideas and inspirations.

The practice of surfing appears to be much more an accepted part of the US journalist's repertoire [4].

Of course, online newspapers, magazines and news services represent just a tiny fraction of the vast files accessible over the Web, and they invariably formed just part of a vast selection of sources exploited. Over half of Internet using journalists said they used 'official' sites generally or named specific institutions from which they obtained data. Government was, predictably, the most heavily visited institution, with the majority indicating that they take material from the various HMG departments, local councils or sites of overseas governments.

University and research sites were nearly as popular as government ones, particularly with specialist and freelance workers for whom, as one pointed out, it saved the time and effort required to 'telephone round, find out who you need to speak to, not catch the person in – or ask for research papers that never arrive'. A good example of their use of academic sites was this one:

> If there was an earthquake I would know what to do ... I would go straight to the USGS site at the University of California. I would go straight to the Global Seismology Unit in Edinburgh ... there is a virtual earthquake site ... a map of all recent earthquakes.

For some specialists the Internet was becoming the only medium by which they

Table 6. Web resources used: sites and site types

Resources	**National and regional journalists**		**Specialist publications and freelancers**		**New Media journalists**		**All journalists**		**Librarians**		**Total**	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Directory (i.e. for address, spellings)	3	14	2	29	0	0	5	12	10	29	15	20
Entertainment/sport	6	27	1	14	3	25	10	24	16	46	26	34
Financial/company	8	36	4	57	1	8	13	32	3	9	16	21
Government	9	41	3	43	4	33	16	39	9	26	25	33
Graphics/pictures	3	14	0	0	1	8	4	10	3	9	7	9
'Institutions' (incl. universities)	11	50	5	71	2	17	18	44	11	31	29	38
News services	5	23	5	71	1	8	11	27	2	6	13	17
Newspapers/ magazines	17	77	4	57	6	50	27	66	13	37	40	53
'Obscure information'	4	16	1	14	0	0	5	12	25	71	30	39
'Official sites'*	17	77	6	86	3	25	26	63	14	40	40	53
Press releases	6	27	3	43	1	8	10	24	2	6	12	16
Science/environment /geography	9	41	4	57	3	25	16	39	9	26	25	33
SDI	2	9	1	4	4	33	7	17	1	3	8	11
Total specifying sites/areas	**22**	**100**	**7**	**100**	**12**	**100**	**41**	**100**	**35**	**100**	**76**	**100**

*Including those naming specific official sites, such as government.

could obtain some material: 'NASA does not issue anything on paper at all. If you are not on email or on the Internet – tough.'

Librarians were on the whole also heavy users of these sites. However, there were differences between the two groups: nearly two-thirds of librarians, as opposed to only a very small minority of journalists, said they used the Internet to procure 'obscure', 'unofficial', or 'offbeat' material. This description included material that was impossible or very difficult to acquire by other means: cult information or strange and unusual facts. Examples included obscure material about the history of Middlesborough in the FA Cup, the Heaven's Gate web site, and the full text of the Unabomber manifesto. The explanation for the difference between librarians and end-users in their use of the Web appears to be that librarians were more able to do these rather more difficult searches. Their greater experience and knowledge of the information retrieval process lead them to experiment. Journalists were generally happy to delegate the more difficult or obscure queries to the librarians.

Another web attraction for media librarians, cited by nearly half of those mentioning specific sites, was entertainment sites, such as BBC television pages, film company promotions and pop star home pages. Librarians also used the Web 'as a glorified telephone directory'. This involved consulting online Yellow Pages, using expert finding services (such as 'Profnet') and confirming overseas telephone numbers.

Communication (email)

Email ought to be attractive to journalists: it is a convenient way of keeping in touch with each other; it is invaluable for overseas communication; it is a means of obtaining information from contacts; and it provides for closer contact with readers. In fact it transpired that the primary use of email was for overseas communication.

Nearly half (48%) of the journalists mentioned the use of email for overseas communication (Table 7). Advantages included speed, price and obviating time zone problems. Obtaining news information via email was far less common, even for foreign correspondents. Journalists based overseas also used email because they felt more in touch with the office. Interestingly this was because of the problems they encountered in phoning in. While the advantages of using email for international communication were generally recognised, emailing within the UK was quite another matter. There was talk from journalists of email being sent 'into the wide blue yonder', of unanswered communication and of the whole uncertainty of whether recipients had read their mail or would reply, and when.

Using email to communicate with contacts was more common amongst the specialist press, freelancers and New Media journalists. Those who did, emphasised the advantages of email over the telephone (if not over personal contact), one saying, for example: 'you can end up having quite chatty relationships with people that might only ever have had the time to talk to you for a few seconds on the phone'. And this can lead to more face-to-face contact rather than less:

> I find myself going out of my way to try and arrange to actually meet people that I have built up this strange electronic relationship with. As time goes by you realise you have not only never seen these people you have never heard their voice.

The extensive use by New Media journalists of email was based as much on cultural reasons as on practical considerations. One such journalist simply could not understand the relatively low take up amongst conventional journalists. For him it created a community ('surely all journalists would want that...'). Another emphasised the democratic qualities of email: it took you straight to the person you wanted to speak to, be it a company executive or the newest trainee recruit: 'with email, everyone is equal'.

For some journalists (13%) email was useful because it kept reporters in touch with the newsroom and vice versa. A Sports Editor, for example, estimated that nearly 50% of his staff were not in the building at any one time. They used email extensively to stay in contact with the office and to file stories – their role

Table 7. Use of email/newsgroups

Uses	**National and regional journalists**		**Specialist and freelancers**		**New Media journalists**		**All journalists**		**Librarians**		**Total**	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Email uses												
For keeping in contact with journalists	7	21	1	9	0	0	8	13	1	3	9	9
For overseas contacts	13	38	7	64	9	56	29	48	16	47	45	47
For communicating with contacts	6	18	6	55	8	50	20	33	2	6	22	23
For communicating with readers (receiving mail only)	8	24	2		5	31	15	25	4	12	19	20
For communicating with readers (receiving and replying)	0	0	1	9	4	25	5	8	0	0	5	5
For contacting Web authors	0	0	1	9	2	13	3	5	7	21	10	11
Other uses	5	15	3	27	1	6	9	15	3	9	12	13
Email attractions/ problems												
Attractions: overcoming time differences	8	24	5	45	8	50	21	34	12	35	33	35
Attractions: masks disabilities/inequality	0	0	0	0	2	13	2	3	0	0	2	2
Attractions: cost	5	15	4	36	0	0	9	15	3	9	12	13
Attractions: facilitates document transfer	1	3	2	18	0	0	3	5	2	6	5	5
Attractions: convenience	5	15	1	9	3	19	9	15	5	15	14	15
Problems: no screening process	5	15	2	18	0	0	7	11	0	0	7	7
Problems: fate of message unknown	5	15	0	0	0	0	5	8	2	6	7	7
Problems: lack of face to face pontaneity	10	29	2	18	0	0	12	20	3	9	15	16
Problems: system failure/incompatibility	8	24	0	0	0	0	8	13	0	0	8	8
Total number using e-mail	34	100	11	100	14	100	61	100	34	100	95	100
Listservs												
Total number using listservs	1	3	3	27	5	31	9	15	8	24	17	18
Newsgroups												
Purpose – story ideas	1		3		2		6		0		6	26
Purpose – finding contacts/experts/ eyewitnesses	4		1		5		10		2		12	52
Purpose – gauge opinion	2		0		2		4		1		5	22
Purpose – technical support	0		1		5		6		0		6	26
Total number using newsgroups	6	18	4	36	7	44	17	28	6	18	23	24

as sports reporters naturally demanding their presence at external locations. The advent of email had led to a reduced presence in the building.

The use of email described above was not typical. There were many doubters and plenty of reasons why journalists have not wholly embraced this potentially important communication channel. Fear of receiving junk email in large quantities represented the most widespread reason why respondents failed to take to email. There were several problems associated with using email for contacts. It was said to lack the spontaneity and flexibility of face to face contact. The latter was highly prized, providing the ill-considered, off the cuff remarks, so often immediately regretted, which was the material of good stories and considerable embarrassment to government ministers, among others. There is also tradition of course. One journalist told us that 'as journalists we are still very conditioned to ringing a human being and asking it and hearing it'.

Communication (Newsgroups)

Paul [8] is one of a number of commentators reporting and recommending extensive use of bulletin boards and newsgroup discussion lists in the United States for consultation and research by journalists. She claimed that they are almost indispensable in, for example, finding people involved in some way in newsworthy situations. Others [9] emphasised the interactive nature of the medium where readers become active agents in the news process. However, one of the most striking results obtained from the research was the poor take up of newsgroup facilities by UK journalists. Just 17 mentioned Newsgroups. The majority (over three-quarters) of respondents either never mentioned this aspect of the Internet or, when asked about it, stated simply that it was not used. A few were frank enough to admit they did not know how to access groups, and others – even some Internet users – did not appear to know what they were. Only a tiny fraction of the sample stated reasons for their non-use, when doing so claiming that the communication is banal, repetitive and of poor quality.

Key information seeking issues

As well as favourable comment the Internet has attracted some criticism too: its potential for overloading users with data, the suspect quality of many sites and the negative impact it will have on much liked traditional information systems and services are commonly mentioned [10].

Information overload

Despite a diversity of comments information overload proved to be a problem for just over half (51%) of journalists (Table 8). Unsurprisingly, perhaps, it was non-users of the Internet who thought it to be more of a problem: 49% of them, but only 28% of Internet users, considered it so. For different reasons 'overload' was not encountered by many novice or experienced users. Those – the majority – not familiar with the fineries of different search engines, advanced searches, directories and subdirectories, etc. happily used the default search engine, took minimal care over query terms and were unconcerned by 20,000 hits. Their usual practice was to scan the first ten, follow a few links and, in the absence of a positive result, turn to a different source.

Advanced users, on the other hand, chose the search engine to fit their query – UK Plus, for example, for strictly British topics, followed directory sub-categories and used Boolean operators or other advanced search options where available. Again, where too many articles were obtained, a few were scrutinised for relevance and the search terminated. There was little sympathy for the overloaded. As one journalist put it 'if you put a search phrase into a search engine and it is not specific enough you get overload, but whose fault is that?' In fact, for some advanced users, the Internet, with its baskets and tailor-made feeds, offered the tools and solutions to overcome information overload.

In the case of Internet users it was often a case of the problem being briefly acknowledged, but then quickly followed by remarks indicating the positive aspects of having access to such volumes of information. Forty per cent of users gave such answers. Surprisingly, too, given the prominence

Table 8. Information overload and the Internet

Overload	National and regional journalists		Specialist and freelancers		New Media journalists		Total journalists		Librarians		Total	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Those considering overload a problem: users	8	20	1	17	2	40	11	28	4	15	15	13
Those considering overload a problem: non-users	17	43	2	33	0	0	19	49	1	4	20	25
All those considering overload as a problem	25	63	3	50	2	40	20	51	5	19	44	30
Those considering overload as not a problem: users	9	23	2	33	3	60	14	36	1	4	15	13
Those not considering overload as a problem: non-users	4	10	1	17	0	0	5	13	5	41	10	9
All those not considering overload as a problem	13	33	3	50	3	60	19	49	12	44	31	26
Why overload isn't a problem:	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Extra material obtained from Internet minimal	4	10	1	17	0	0	5	13	2	7	7	6
Worth hunting through Internet to find valuable information	2	5	1	17	2	40	5	13	2	7	7	6
Quantity is good in itself	4	10	0	0	1	20	5	13	5	19	10	9
Internet is unobtrusive	1	3	1	17	0	0	2	5	2	7	4	3
Accessibility of obscure documents	1	3	0	0	0	0	1	3	1	4	2	2
Why overload is a problem	***n***	%	***n***	%	***n***	%	***n***	%	***n***	%	***n***	%
Time taken to retrieve relevant information	7	18	2	33	0	0	9	23	3	11	12	10
Poor search engines	3	8	1	17	0	0	4	10	3	11	7	6
Too much information	5	13	0	0	0	0	5	13	5	19	10	9
Poor quality	4	10	0	0	0	0	4	10	2	7	6	5
Wider issues (i.e. too much information in society)	4	10	4	67	2	40	9	23	1	4	10	9
Solutions to overload	***n***	%	***n***	%	***n***	%	***n***	%	***n***	%	***n***	%
Selectivity (i.e. only Government sites)	17	43	6	100	3	60	26	67	1	4	27	23
Use of bookmarks	10	25	3	50	2	40	15	38	14	52	29	25
Adoption of time limit	3	8	2	33	0	0	5	13	12	44	19	38
Limiting hyperlinking	2	5	0	0	0	0	2	5	7	26	9	16
Advanced search strategies	2	5	2	33	0	0	4	10	5	19	9	16
Total mentioning overload: users	22	55	3	50	5	100	25	64	12	44	67	57
Total mentioning overload: non-users	18	45	3	50	0	0	14	36	15	56	50	43
Total mentioning overload	40	100	6	100	5	100	39	100	27	100	117	100

*In all cases percentage figures are of the total in each category mentioning overload specifically.

N.B. Some respondents gave more than one example of overload not being a problem; and others contradicted themselves, first saying it was a problem but then insisting this did not matter. Therefore figures have to be treated with some caution.

accorded the issue of overload in the professional information literature, librarians were less worried about overload than journalists – just 19% considered it a problem. Some dismissed overload on the basis that it was all part of the librarian's role to confront and solve the problem.

Where journalists were unconcerned about overload it was because wallowing in information was nothing new and it was something that they were quite prepared for. More than one journalist claimed that you could not have too much information. Indeed, the quantity of information was actually cause for celebration. There were other reasons why information overload was not a cause for concern:

- The extra material yielded from the Internet was relatively small compared to what was already available. In the words of one journalist: 'any one of the Sunday papers contains enough words to provide almost the complete text of Dickens. *The Guardian* every day is the equivalent of *Our Mutual Friend*. ... But we read *The Telegraph* and *The Times* and we listen to The Today Programme[4]... You throw in the Web – that's nothing.'
- It was worth hunting through trivia to find valuable information. A journalist explained: 'Searching the Internet is a bit like the toy shop Hamleys – if you want a dinky toy which is on the third floor ... but en route there are various goodies which side-track you.'
- The Internet is unobtrusive: 'You can sit it in a corner and tell it not to disturb you and it won't' said one journalist.
- The ease with which you could access documents on the Internet more than compensated for any problems associated with finding too many of them. Thus a journalist told us that 'the time that it might take you to get in touch with a research unit of a university somewhere and get hold of a particular item could go into days or even weeks. But (you can) go to the Web and find their page and get it in seconds.'

Non-users were not so easily convinced. Their complaints were not only levelled at the excess of information available but also at the Internet search engines; and there were also the worries about the time that it would take to sort and filter the documents retrieved.

Quality

The majority of journalists were phlegmatic about the quality of data they obtained on the Internet and the problem of verifying it. It concerned, rather than worried them, but no more than information obtained from other sources. The particular problem of authenticating email was especially acknowledged. Whether quality was seen to be an issue depended on three factors:

1. job role and the different information needs that arose;
2. experience and knowledge of the Internet;
3. attitude towards the nature of journalism and what constituted a newspaper.

Job role and information needs

Journalists perform many roles, and hold a variety of posts. Indeed, a good number of journalists have nothing to do with the reporting of news. They may be involved in criticism, comment, analysis, the obtaining of gossip, management or other tasks that may require information of various types in various quantities – from hard facts to whimsical speculation, from government reports to innuendo. Clearly, then, the quality/authenticity question is tied in with the nature and type of information required in carrying out one's job. Editors figured strongly amongst those who appeared to worry most about the authenticity of data, having, of course, responsibility for the reputation of their papers.

Those whose subject fields required particular attention to the veracity of information – crime, business and science are cases in point – were also cognisant of the problems. Some had concerns peculiar to the type of data they handled. As a leading crime correspondent explained:

> There is a problem with inaccurate information and particularly in my field that could be very dangerous because we run into problems of libel and con-

---

[4] A UK current affairs radio programme.

> tempt. Reporting crime if you get bogus information or inaccurate information about people's convictions or about crimes then you are in trouble and I am way off the Internet for that reason.

Of course not everybody is seeking the truth. There is a high demand for the controversial, gossip, etc. Thus the fact that the Internet contains information of doubtful, but interesting, pedigree can be seen as a journalistic plus. In particular features journalists and those charged with producing articles of unusual 'human interest' tended to be less preoccupied with the quality of information.

The nature of the information required (and the source of the query) also shaped librarians' approach to authority and the Internet. Nevertheless, fewer librarians than journalists were worried about the quality issue, and there was a general feeling that it was the library staff who had to counsel the journalists regarding this matter. They were taking on their professional shoulders the corporate responsibility for accuracy and quality.

Experience and knowledge of the Internet
Not surprisingly those journalists well experienced in using the Internet were more aware of its limitations. Used to being fed on a diet of wire and other online services considered to be authoritative, many of the journalists who made little use of the Internet had clearly not considered the issue with regard to the Internet.

The nature and role of newspapers
Newspapers are not academic journals and thus do not require the same levels of proof and referencing. One information manager explained that there is a marked contrast between, for example, medical research and newspaper reporting. Someone engaged on the former, investigating the effect of a certain drug is required to consult and evaluate all appropriate literature and proceed with due consideration of all the evidence documented therein. A journalist researching the same thing may be happy to read two or three reports, and interview a few patients or users. There simply is not the same onus on the reporter to apply the same scientific rigour to his research as a doctor might.

Verification strategies
In general the strategy was simple: those seeking facts tend to use sites of authoritative bodies. There were other strategies employed:

Double checking: Few journalists used the Internet in isolation. Many journalists either only used the Internet where the information was already known to them in a general way and confirmation or additional details were sought or, alternatively, as a first step for some background information researched in more detail later from other sources.

Using subject expertise: Some journalists were untroubled by the authority issue by virtue of their familiarity with their subject or their contacts. Particularly specialist journalists mentioned this. For example, one interviewee heavily involved in Middle Eastern affairs was prepared to use communiqués posted onto the Internet by disparate political and religious groups. He knows their aspirations, beliefs and practices and the way they articulate these. It would be no problem for him to identify false association with any particular faction.

Passing the buck: As one journalist explained: we state explicitly in our articles that 'the information came from the Web: that way we are covered'.

Displacement of other information/ communication sources

No other communication medium or information system seems to have been diminished to any great extent as a result of Internet use. More often than not Internet use seems to stimulate information seeking in general, resulting in *increased* use of other sources. A New Media journalist explained he reads more books because of the Web – he has found some sites of excellent book reviews, and is thus persuaded to buy more books. Take the case of the telephone, too. Internet users rarely phone the press office of a company for a copy of a press release or annual report; instead they visit the company's home page and download the document – it is easier to process it that way, and as one journalist said: 'the press office is always open'. During non-business hours, when sources are not available, web sites are playing an important role in delivering information to the Media. In this case the Internet has resulted in a decline in phone calls. However, the following up of this data is undertaken over the phone, thus increasing its use.

That there was little or no displacement occurring in the traditional media companies might not be surprising – after all, only a minority of journalists are Internet users. However even amongst Internet users there is little displacement. There are good reasons for this:

- The Internet filled the information gaps left by other sources. At Time Life, for instance, the sites searched on the Internet do not hold the sort of data stored in the past.
- There is a general mistrust of the Internet, both in terms of information continuing to be posted and whether such information will remain free. An excellent example is that of *Hansard*, available originally in hard copy, later on CD-Rom, and currently also on the Internet. Although at least three organisations used *Hansard* on the Internet they still retained hard copy access.

Where there is evidence of real displacement is in the case of overseas newspapers, which are now being purchased less, with titles such as *The Washington Post* available on the Net. Commercial online services also appear to be at some risk. Various interviewees speculated that the presence on the Internet of, for example, the full text of *The New York Times*, various regional US titles, *The Daily Telegraph* archives etc. must make a difference to online sooner or later. At *The Guardian* online use of FT Profile has tumbled with bills noticeably dropping over the last months of the project.

Key findings

Take-up and impact

Journalists, far from being in the vanguard of the Internet revolution, were nothing more than followers – (partly) held up by editorial systems – once the envy of many, but now sadly out of date. The situation varied throughout the Media – and is changing as we write – but in general it is unlikely that more than one in five journalists use the Internet[5]. Lack of access, the richness of existing information provision, suspicion of the new resource (and its disciples), and a shortage of time were the contributing factors. There did, however, appear to be a general feeling that it is only a matter of time before most journalists become Internet users, although this was expressed without real enthusiasm. There was also little evidence to suggest that those currently using the system would extend or improve their searching skills. Factors conspiring against dramatic increases in Internet use were that journalists did not feel it was their job to undertake online searches, and they would not have time to learn to search the Internet efficiently.

The early leaders

If the lack of Internet take up in the workplace was unexpected, another surprise was the characteristics of those who have taken to the Internet. Far from them being the stereotypical young and male, most are well practised journalists well into their thirties/forties[6], and in some media companies women outnumber men. One reason is that the people of these age groups, having emerged successfully from the scramble to get established in their early careers, can experiment with the Internet at work and have the resources to indulge in home PCs with Internet access. Another reason, is that the 'older hands' who saw the introduction of online ten to twelve years ago, assimilated it into their journalistic practices and were now viewing the Internet in much the same light. There is however one area of the Media where Internet users do conform to type – and that is in the New Media companies (VirginNet and *The Guardian* online). Here journalists were young and male.

---

[5] The number of Internet users in the UK in 1997 was estimated to be 10% of the population [11].

[6] This finding is supported by *The Times* [12] which found that: 'most web users are in the 35–45 age range'.

Uses

'Visionary' CAR journalists/librarians in the USA describe posting questions on newsgroup lists, email facilitated real time interviews, investigative reporting from the Web and much more. We see few signs of such innovative and adventurous Internet use here in the UK. Where the Internet was being used, it was being used in a very limited way. When given the powerful, comprehensive, freely available ultimate end user tool (the Internet), the tendency was to stick with what you knew, what was familiar (newspapers and official sites – something which Garrison [3] also found) and do just the simple things (fact checking, background information). In many respects the Internet was used as a giant cuttings collection.

Libraries and their position

The arrival of every new technology is normally greeted with a flurry of articles writing the obituaries of the library and librarian – and it is interesting to learn that similar things are beginning to be said about newspapers and journalists [13]. There was no sign of this happening. In particular, the situation with librarians could hardly be more different. For this group Internet use tended to be universal and heavy. Hardly any of them felt that the new technology was a threat to their jobs – more a case of more work and greater opportunities. This appeared to be another instance – full text databases was the last one – where librarians have employed and pioneered the new end-user technologies to their own advantage.

Impact on other information systems

If not libraries then possibly some other information/communication system is going to be elbowed aside by the Internet? It seems not – not yet, anyway. No other communication medium or information system has greatly lost out as a result of growing Internet use. Web sites stimulate the greater use of other sources. However, the situation does appear precarious for the commercial online host. There is increasing evidence that thrifty information professionals are searching the Internet's free newspapers before going to FT Profile – and some newspapers are already seeing some of their online bills plummet by 50% as a direct result. Most worryingly for the hosts, new markets, like freelancers and contractors, a rapidly expanding group, blanch at the very prospect of paying what they see as an unacceptable level of online costs. They are increasingly drawn to the Net because they see it as a free online and library service.

More often than not though the Internet *both* reduces and increases the use of a source leaving usage levels at what they were, but having slightly changed the nature of use. This is particularly the case with the telephone. Email has yet to take the Media by storm, but when it does – as it surely will – the day of the fax must be numbered.

The Internet issues

The authority and permanence of much web data has been questioned, the volume of data available and even its impact on the traditional sources of information merits unfavourable comment. It is difficult to be sure, because non-users were under-represented in our study – and they might well have had contrasting opinions – but only half of all journalists were concerned with these issues. Many journalists were also phlegmatic about the poor quality of some of the data obtained on the Internet and the problem of verifying it. It did not especially worry them because it was felt that the good journalist instinctively doubts the quality of all information, is trained to 'second source' everything, and therefore sees the Net as another source to be checked. Journalists have a 'feel' for the authenticity of the data, judging it from long years of experience. Their judgement is no less prone to error, in dealing with, say, telephone calls than it is with the Internet.

Conclusions

Few journalists in the UK used the Internet, but even fewer had made up their minds about it. The Internet and the issues it confronts are so enormous that this is probably inevitable. For most of the people we spoke to, apart from New Media journalists, the

Internet was nothing revolutionary and certainly not a panacea for all their communication and retrieval problems, of which they had preciously few. Nevertheless, there was a palpable sense that major changes were in the air and that they were just a few steps down a potentially long road of discovery. The real question is whether the changes in structure to the industry, occasioned by the migration of news from hard copy to the Web (a process still in its infancy), will see the New Media model of information seeking being adopted by significant numbers of traditional journalists. In many ways the Internet in the UK newsroom is, then, if not a sleeping giant, certainly one whose faculties have not all been awakened yet. It has definitely arrived, and with more than a whimper – but not (yet) with the expected big bang. Indeed, it is very unlikely that there will be any dramatic changes, just a slow absorption of the technology. The respected media commentator, Simon Jenkins [14], was more right than wrong when he proclaimed the Internet to be just a niche service, but he is surely wrong when he says that it will 'strut the stage for only a short time'. All technologies are talked up – by the purveyors of hardware and software and by information professionals looking to enhance their job. The reality is that impacts are always slower and gentler than envisaged and that existing communication and retrieval technologies refuse to fade away – and are in many cases re-energised by the new, incoming technology.

The situation portrayed in the US literature – albeit painted largely by web and mailed questionnaires to self-selecting samples, could not be more different. However, there would appear to be very good reasons for this. Firstly, most journalists in the UK, unlike their US counterparts, do not have desktop access to the Internet. Secondly, journalism practises differ: generally newspapers in the UK would not allow a story to be sourced entirely on the Web, and there is a definite preference for primary rather than secondary data – of course, some web data is primary. This, together with the absence of much government data on the Web, explains why computer aided reporting has failed to take hold in the UK.

References

1. Nicholas, D., Williams, P., Cole, P. and Martin, H. *The Media and the Internet*. London: Aslib, 1998.
2. Nicholas, D., Harris, K., and Erbach, G. *Online searching: its impact on information users.* London: Mansell, 1987.
3. Garrison, B. *The Web in the newsroom 1997.* 1997. http://www.miami.edu/com./car/spj-97/index.htm (visited 8 April 1998).
4. Middleberg, D. and Ross, S. *The Media in cyberspace IV: a national survey.* 1998. http://www.mediasource.com/study/CH01.HTM (visited 2 April 1998).
5. Paul, N. *Computer assisted journalism: the Four Rs of CAJ.* 1996. http://www.facsnet.org/report_tools/CAR/carfourr.htm (visited 8 April 1997).
6. Nicholas, D. The information needs interview: a long way from library-use statistics. *Education for Information*, 15(4), 1997, 343–350.
7. Hancock-Beaulieu, M. *Evaluation of online catalogues: an assessment of methods.* London: British Library Research and Development Department, 1990.
8. Paul, N. *Computer assisted research: a guide to tapping online information.* 3rd ed. St Petersburg: Poynter Institute, 1997.
9. Outing, S. A decade of online newspapers – lessons and trends. *NetMedia 97 Proceedings.* London: City University, 1997.
10. Coy, P., Hof, R.D. and Judge, P.C. Has the Net finally reached the wall? *Business Week*, 26 August 1996. Also available at: http;//www.businessweek.com/1996/35/b3490107.htm (visited 12 December 1997).
11. *NUA Internet Surveys.* 1997. http://www.nua.ie (visited 20 May 1997).
12. The Times Interface. Everyone knows about the net. *The Times Interface*, 23 April 1997, 2.
13. Nicholas, D. and Frossling, D. The information handler in the digital age. *Managing Information*, 3(7/8), July/August 1996, 31–34.
14. Jenkins, S. No plugs, no wires, no rivals. *The Times*, 4 January 1997, 16.

Citizenship information needs in the UK: results of a national survey of the general public by personal doorstep interview

Rita Marcella and Graeme Baxter
School of Information and Media, The Robert Gordon University, Faculty of Management Building, Garthdee Road, Aberdeen AB10 7QE
r.marcella@rgu.ac.uk

This paper reports the results of the second stage of the Citizenship Information research project funded by the BLR&IC: a national survey, by personal doorstep interview, of the citizenship information needs of 898 members of the UK public. Major findings include: that the public obtain most of their information on current issues via the media, and that they generally feel well informed on these issues. The public believe, however, that government is not doing enough to inform them about the Single European Currency and local council cutbacks. Small proportions of the sample had encountered problems in relation to employment, education, housing or welfare benefits and had consulted a range of information sources in order to solve these problems. Over a quarter of respondents had experienced disadvantage through a lack of access to information. Significant proportions of respondents were poorly informed about legal rights, welfare benefits and local politics. A highly significant majority (91.7%) believed that freedom of information was important for exercising their rights as citizens. Although access to computers in the home is presently limited, the majority of respondents indicated a willingness to use computers to vote and interact with government. Public libraries were the preferred source of government information and were seen as appropriate locations for a range of other types of citizenship information.

Introduction and methodology

The April 1999 issue of *Aslib Proceedings* contained a report [1] of the results of the first stage of the Citizenship Information research project funded by the British Library Research and Innovation Centre, a questionnaire-based survey of a sample of the UK population designed to elicit preliminary data on their use of and need for citizenship information. These questionnaires had been distributed in public libraries, Citizens Advice Bureaux (CABx) and other generalist information and advice agencies throughout the 13 Government Office Regions of the UK.

This current paper reports the results of the second stage of the project: a complementary national survey of the public's citizenship information needs, this time by personal doorstep interview. These interviews were designed to elicit more qualitative, extended and individual responses, and to form a more in-depth and less pre-determined set of data. In conducting the interviews at people's homes, away from the 'institutional' setting of the initial survey, it was expected that any deficiencies in the first survey would be overcome by reaching, for example, the less literate respondent and those from lower social grades.

To conduct the interviews, the project team employed undergraduate and postgraduate librarianship and information studies students from library schools throughout the country. Interviews took place in 12 of the 13 UK regions (the exception being Merseyside) and in the same towns and cities in which the Stage 1 questionnaires had been disseminated. Each interviewer was provided with socio-economic and population data for their allocated town/city taken from the *1991 UK Census Small Area Statistics* [2], and was asked to identify five local government electoral wards in which the survey would take place: the ward containing the highest percentage of residents belonging to *Standard Occupational Classification* [3] Social Classes IV and V (partly skilled and unskilled occupations), the ward with the lowest percentage of Classes IV and V, and three other wards spread across the cumulative population figures for the town/city.

Each interviewer was asked to complete 15 interviews in each of the 5 wards – a total of 75 interviews in each town/city. All were successful, with the exception of one interviewer who fell two short of the desired quota; therefore 898 interviews were completed. The interviews took place between May and November 1998. The interviewers were instructed to use the random walk sample method, whereby they randomly selected a starting point within each ward and proceeded to attempt an interview at every seventh household within that ward until the required quota had been reached. Detailed instructions were provided on how to deal with road junctions, vacant households, blocks of flats, offices and commercial premises, etc. The response rate of this method in individual wards ranged from 31.9% (i.e. one successful interview obtained at every third household visited) to 4.9% (one in every twenty homes visited).

Table 1 illustrates the gender and age group of the respondents. The female-male ratio was significantly greater than that occurring nationally (51%-49% [4]) and in the first survey. The proportion of older respondents (40% aged 55 or over) was also substantially greater than the national percentage of 26% [5] and that in the initial survey.

In total, 56 (6.2%) of the respondents belonged to an ethnic minority group, while 118 (13.1%) indicated that they were disabled in some way. 16.7% lived in rural areas. Just over a quarter had completed an undergraduate or higher degree course, almost 23% had completed a further education course, while just over half had received no education beyond school. Over 56% of the respondents were economically inactive (i.e. student, retired, running a home or permanently unable to work), slightly greater than the national figure of 51.1% [6]; however, as the national figure includes all people under 16, who were excluded from this survey, then it becomes clear that the respondents came to a greater extent proportionally from the economically inactive groups than for the UK as a whole. These data would suggest that the doorstep interview methodology has reached greater proportions of those groups deemed to be in danger of social exclusion, providing results that are particularly revealing. They also complement those of the first survey where a more dynamic and arguably more 'informed' group of respondents was achieved.

Survey results

Respondents were initially asked to identify what they believed to be the single most

Table 1. Gender and age of respondents

Age group	Male	Female	Not specified	Totals	
				n	%
16–19	18	24	–	42	4.7
20–29	72	72	–	144	16.0
30–44	75	149	–	224	24.9
45–54	40	82	1	123	13.7
55–64	54	77	1	132	14.7
65–74	52	75	–	127	14.1
75 or over	35	66	–	101	11.2
Age not specified	1	4	–	5	0.6
Totals	**347** **(38.6%)**	**549** **(61.1%)**	**2** **(0.2%)**	**898**	**100.0**

important issue in UK current affairs at the time of the survey. A wide range of issues was cited, although the top ten responses were: the Northern Ireland situation; the state of the education system; health issues and the National Health Service; the European Union and UK participation in European Monetary Union; the UK economy; unemployment; law and order; the football World Cup; the environment; and the state of the Labour Government. There were some significant demographic differences in the response to this question. When compared with female respondents, males placed greater importance on European issues (+6.0 percentage points) and unemployment (+4.3), but less importance on health (–8.6) and education (–5.4). Younger respondents appeared to have less interest in health and (perhaps surprisingly) Europe but more interest in education. Amongst the better-educated respondents there was more interest in economic and environmental issues but less interest in the state of the NHS.

Respondents were asked from which source they had obtained most of their information on their chosen current affairs issue. The media dominated overall for subjects, with at least one form (television, radio or newspapers) being cited by over 75% of respondents in each case. Use of television as the main information source was greater among the manual and unskilled workers, while use of newspapers was greater amongst those with professional and managerial occupations.

Most respondents felt well or very well informed on the main issues identified, although over 15% in each case felt poorly informed about Europe, Northern Ireland, health, law and order, and the environment. Women tended to feel less well informed than men; while those who had received a better education felt better informed than the other respondents. Despite feeling well informed, significant proportions (over a third in each case) wanted to find out more about education, Europe, the environment, health, unemployment, the economy, Northern Ireland and law and order. When asked where they would most likely go to obtain further information, high proportions indicated that they would approach public libraries. Indeed, libraries were seen as the most important single resource for the environment (46.7%), education (38.8%) and health (29.3%), and an important source of information on employment, Europe, law and order and Northern Ireland.

Respondents were asked if their chosen issue significantly affected their own lives. While the economy (76.7%), health (69.6%), the environment (69.6%), education (67%), Europe (60.5%), unemployment (57.9%), and law and order (54.5%) were deemed by the majority of respondents to have an impact on their lives, the other topics were felt to be less central. However, of the Northern Ireland (i.e. Belfast) respondents, 66.7% felt the local political situation significantly affected their own lives.

The next part of the survey consisted of a more detailed investigation of three current issues chosen by the researchers as representative of citizenship information need at international, national and local levels: the Single European Currency, the BSE crisis and local government financial cutbacks.

Just under 75% of respondents indicated that they would like to find out more about European Monetary Union, with female respondents displaying a greater interest than males (+10.2 points). The mass media (38.7%) and public libraries (21.7%) would be the most popular sources of information, followed by a government information programme (8.7%). The vast majority (80.4%) felt that the government was not doing enough to inform the public about the Single Currency.

The majority of respondents (59.2%) felt that they were well informed about BSE. The media (26.6%) and public libraries (19.4%) were again the most likely sources for those who felt poorly informed, with doctors and other health professionals (7.9%) forming the major alternative source, particularly for women and the less well educated. A slight majority (55.2%) felt that the government was not doing enough to inform the public about BSE. When asked how they might make their own feelings known to the government on this issue, the preferred methods were contacting a local MP (38.6%) or a government minister or department (10%), although a large proportion (35.6%) were not interested in making their own feelings known.

Almost two-thirds (63.4%) of respondents felt poorly informed about local government cutbacks, with the most likely sources of additional information being local council offices (31.1%), local newspapers (11.6%) and public libraries (10.4%). The majority (71.6%) believed that their local council was not doing enough to inform them about cutbacks. In making their own feelings known on this issue, significant proportions would contact local councillors (20.3%), their local MP (19%) and local council offices (17.5%), although over a quarter (26.2%) were again not interested in making their feelings known.

The next part of the survey focused on what might be termed survival information, i.e. information that might help individuals to overcome the problems that occur in day-to-day life. It asked whether or not respondents had encountered a problem relating to employment, education, housing or welfare benefits, and elicited brief details on the nature of these problems plus the information sources used to help solve the problems. In total, 328 (36.5%) of the respondents indicated that they had had a problem relating to one or more of the four issues (see Table 2).

141 respondents (15.7%) indicated that they had had an employment problem, including finding work (the most common problem, cited by 71 respondents), dealing with redundancy or unfair dismissal, and sexual harassment or bullying in the workplace. Job Centres and newspapers were the information sources used most frequently to help respondents solve employment problems. Interestingly, employment problems were cited by a greater proportion of males (20.2%) than females (12.9%).

Table 2. Type of problem encountered

Type of problem	*n*	%
Employment problem	141	15.7
Education problem	118	13.1
Housing problem	122	13.6
Welfare benefits problem	135	15.0
Total who cited one or more of the above problems	328	36.5

A smaller proportion (118, 13.1%) cited an education problem, such as selecting a school or university, obtaining funding and grants, and special needs or health problems affecting education. Local education authorities, schools and universities, and family and friends were the most frequently cited information sources. Unsurprisingly, education problems were cited more frequently by students and those aged 16–44.

Housing problems were indicated by 122 (13.6%) respondents; these included obtaining council housing, dealing with problem neighbours, and homelessness. Most of these respondents had approached council housing departments and local councillors for help. Housing problems were cited most frequently by those aged 20–44.

Exactly 15% (135) of the respondents had had a welfare benefits problem. All were concerned with claiming or establishing entitlement to a wide range of benefits. The main sources of their information were Department of Social Security (DSS) Offices, CABx, council social services departments, and family and friends. A greater proportion of rural respondents (22.7%) cited a benefits problem than did the urban respondents (13.5%); while disabled respondents had encountered more problems than able-bodied respondents (+11.9 points).

The survey then explored the concept that exclusion from information creates disadvantage. Respondents were asked to give an example of an occasion when not being able to obtain information had proven a disadvantage and over a quarter (26.1%) provided such an example. These related to a wide range of issues, including financial matters, consumer issues, and health and relationship problems. Interestingly, examples were cited more frequently by those respondents that might not normally be associated with disadvantage, i.e. those from higher social classes and those with a higher level of educational qualification. This might relate more to a lack of 'awareness' of disadvantage amongst those in the lower social classes and the less well-educated.

The next section of the survey sought to measure how well informed the respondents were about citizenship information matters, and to examine their levels of participation in the democratic process. They were firstly asked how well informed they felt they were about five different citizenship topics: national politics, local politics, legal rights, welfare benefits and equal rights (see Table 3). The majority (i.e. between 50% and 70%) felt well or adequately informed about these subjects, although over a quarter of the sample in each case felt poorly informed about legal rights, welfare benefits and local politics. Male respondents felt better informed than females about national politics (+14.7 points); while with regard to welfare benefits, disabled respondents were better informed than able-bodied respondents (+16.9).

Respondents were provided with a list of six types of group or association and asked if they were a member of any. Overall there were low levels of participation in these groups, with the highest proportions being for membership of leisure (32.4%) and charitable (25.2%) groups. Only small minorities were members of professional and trade associations (16.6%), trade unions (13.6%), pressure groups (6.9%) or political parties (6.7%). Membership of all types of group was higher amongst the better-educated respondents.

Respondents were asked how regularly they vote at elections. The numbers who said they always vote at national (77.2%), local (65.8%), and European (45.3%) elections were higher than the most recent actual turnouts for these votes [7, 8, 9] particularly for local elections, suggesting that respondents tend to overestimate their participation in this respect. In general, the proportions who always vote at all three types of election grew larger as the respondents grew older. The survey also asked respondents whether, apart from voting, they actively participated in the political process. Only a small proportion (10.2%) participated in other forms of political activity, such as canvassing, campaigning, and attending political meetings. Participation was generally higher amongst the professional and managerial occupations, and those who had received a more advanced education.

The survey asked respondents for their opinion on the importance of freedom of information, and if they felt that access to accurate and unbiased information was important for exercising their rights as citizens. The vast majority (91.7% in both cases) felt that these issues were important, reinforcing the finding in the first survey that indicated that 94.3% of respondents believed that information access is important for exercising their rights. Respondents were also asked to explain why they had responded in the manner they did. In summary, freedom of information was felt to be: a basic democratic right; necessary to ensure politicians' accountability; and necessary in the current atmosphere of sleaze and corruption. Access to personal data was deemed a particularly significant issue. Respondents felt that access to accurate and unbiased information was essential to their capacity to make effective decisions or form opinions.

The next section of the survey paid particular attention to the use of Information and Communication Technologies (ICTs). In a closed question, respondents were asked

Table 3. Respondents' awareness of various citizenship topics (%)

Topic	Expert	Well informed	Adequately informed	Poorly informed	Not interested
National politics	1.0	24.2	44.1	13.9	16.5
Local politics	0.6	16.5	40.5	27.2	15.0
Legal rights	1.1	17.7	40.9	31.3	8.6
Welfare benefits	2.1	17.1	31.8	30.7	17.8
Equal rights	2.6	22.0	41.8	20.7	12.5

about current levels of access to ICTs. Access to computers (39.3%), the Internet (14.6%) and satellite/cable television (26.4%) in the home was limited, while only 12.6% (35.3% of those in employment) had access to the Internet at work. Unsurprisingly, access to ICTs tended to decrease as respondents grew older.

The respondents were asked if they would use computers to take part in any of four different democratic activities, if computers were available for them to use either at home, at work or in public places. Table 4 summarises the response. A slight majority in each case would be willing to use computers to vote, make their own feelings known or obtain government information, but there was a reluctance to use computers to participate in debates. In general, those respondents with a higher occupational classification and/or a higher academic qualification were more willing to use a computer for these purposes. Older people, people in rural areas and people with a disability were less willing to use computers than younger respondents, urban residents and able-bodied respondents. These findings reinforce concerns over specific groups deemed to be in danger of exclusion by increasing emphasis on ICTs in communications with government.

Respondents were asked if they might have concerns over confidentiality if they were to use a computer to interact with government. Of those respondents who would use a computer to vote, 47.7% were worried about confidentiality. Similar figures were found amongst those who would use computers to debate (47.5%) and provide feedback to government on particular issues (50.6%).

From a list of five options, respondents were asked how they would prefer to make their own feelings known on a subject they felt strongly about. The majority of respondents (66.4%) preferred the traditional method of writing to someone, with smaller proportions preferring to speak at a public meeting (5.8%), debate by computer (5.5%), telephone in to a radio/TV debate (4.7%), speak on a televised debate (2.3%). A significant proportion of 15% were not interested in making their feelings known publicly.

The survey then turned its attention to the public's use of libraries and other information sources. From a list of 11 organisations, individuals and locations, respondents were asked from which one they would prefer to obtain official government information. The results are illustrated in Table 5. Public libraries were the most popular source, cited by over 43% of respondents. This is a significant reinforcement of the finding in the first survey, where 77.3% of respondents indicated that they would approach libraries for information on at least an occasional basis. Interestingly, disabled respondents, when compared with the able-bodied, had a preference for approaching CABx (+15.4 points) but were less inclined to visit a library (–13.7).

The respondents were also asked for what reasons they would prefer to obtain information from their chosen source. For most respondents it was the accessibility and the fact that they visited it regularly that made the library their preference; libraries were also seen as reputable and having an appropriate atmosphere. Those that preferred MPs and government departments were motivated by getting their information directly from the primary source. CABx were felt to have

Table 4. Potential use of computers for democratic activities

Activity	**Yes %**	**No %**	**Don't know %**
To vote in elections	54.7	44.2	1.1
To take part in political debates	35.0	63.5	1.5
To let government know what you think about particular issues and decisions	53.5	46.1	0.4
To obtain official government information	59.8	39.5	0.6

Table 5. Preferred source of Government information

Source	*n*	%
Public libraries	388	43.2
Post offices	132	14.7
Citizens Advice Bureaux	108	12.0
Local MP	52	5.8
Local council offices	49	5.5
Offices of government departments and agencies	36	4.0
Schools	32	3.6
A computer in a public place, such as a shopping centre	31	3.5
Banks and building societies	18	2.0
Medical centres and doctors' surgeries	9	1.0
Community and leisure centres	4	0.4

very helpful staff (libraries did not rate highly here) and were also seen as particularly reputable. Overall, the primary motivating factors appeared to be accessibility and frequency of visit.

72.5% of the respondents were members of a public library; this is higher than the recently quoted national figure of 58% [10], and interesting given the survey methodology and demographics of the sample. Membership was higher amongst the higher social classes and the better-educated and was lowest amongst those seeking work (64.6%) or running a home (63.9%).

Respondents were asked if they felt that public libraries were suitable places for obtaining information on the following five topics: their local council, welfare benefits, job opportunities and careers, housing, and consumer advice. Their response is illustrated in Table 6. The vast majority (82%) believed that libraries were appropriate locations for council information, while smaller majorities in each case felt that libraries should hold information on the other four topics. There appeared to be some doubts over whether libraries should hold housing information, with 38.2% of respondents believing libraries to be unsuitable locations for such material.

Those respondents who believed libraries were not appropriate places for the five types of information were asked to suggest more suitable locations. Local council offices were regarded as appropriate locations for council information; CABx, DSS offices and council social services departments were felt to be more suitable for benefits information; it was felt that jobs and career information should be located in job centres; council housing departments were deemed more appropriate for housing information; while CABx and Consumer Protection or Trading Standards offices were believed to be suitable for consumer information.

The survey then investigated the respondents' awareness of the types of information that public libraries hold. They were asked if they were aware that most libraries hold official government publications and information on health care, educational opportunities, legal rights and the European Union. Table 7 illustrates the response. Overall, 74.8% of respondents were unaware of at least one of the information types held by libraries, pro-

Table 6. Suitability of public libraries for obtaining information on particular topics

Topic	Suitable %	Not suitable %	Don't know %	Not interested %
Your local council	82.0	11.7	4.8	1.6
Welfare benefits	65.1	27.4	5.4	2.0
Job opportunities and careers	61.4	30.7	6.1	1.8
Housing	53.5	38.2	6.3	2.0
Consumer advice	67.8	23.6	6.5	2.1

viding evidence of a real need for libraries to promote their collections more effectively. As can be seen, a small number of respondents stated that, while they did not actually know that libraries held this information, they would have presumed that they did so. There was less awareness of libraries holding government publications amongst the lower occupational classes and, with the exception of health care information, there was a greater awareness of library holdings amongst the better-educated respondents.

Finally, those respondents who were unaware of at least one of the information types held by libraries were asked how libraries could let them know more about the types of information they hold. Publicity campaigns, leaflets and posters, exhibitions and displays, improved signposting and mailshots were all recommended. Unsurprisingly, non-members tended to emphasise the need for external publicity and mailshots, while current library members believed that there should be more promotion of materials in the library. Some respondents felt that libraries should not waste their resources by undertaking costly promotional activities.

Table 7. Awareness that particular types of information are held by most public libraries

Type of information	Aware %	Unaware %	Presumed %
Official government publications	53.5	45.8	0.4
Health care	53.6	45.1	1.0
Educational opportunities	69.0	30.2	0.4
Legal rights	53.7	44.9	1.1
European Union	36.9	61.1	1.7

Conclusions

The survey response (898 interviewees) was representative in most respects, but held greater proportions of women, the elderly, and the economically inactive than figures for the UK population as a whole. It is believed that high representation amongst these groups, who may face social exclusion, reveals particularly interesting results.

Significant current issues identified by respondents were dominated by Northern Ireland, education, health and Europe. The media was the prime source of information on most current issues, but for those respondents who wished to obtain further information public libraries would play a prominent role.

While the majority of respondents felt well informed about the BSE crisis, they would like to find out more about the Single European Currency and local government cutbacks. The majority of respondents believed government should be doing more to provide information on all three topics. In terms of survival information, small but significant proportions of the sample had encountered employment (15%), education (13.1%), housing (13.6%) or welfare benefits (15%) problems, and had utilised a variety of information sources and providers in order to solve these problems. Over a quarter of respondents cited an example of an occasion when a lack of information had proven a disadvantage.

The majority of respondents felt well informed about areas relating to citizenship, although more than a quarter were poorly informed in legal rights, welfare benefits and local politics. There were low levels of participation in groups and societies. While respondents tended to overestimate their voting patterns, only small numbers participated in other forms of political activity, such as canvassing or campaigning.

A highly significant majority (91.7%) of respondents felt that freedom of information and access to accurate and unbiased information were important in exercising their rights as citizens.

Access to ICTs is currently limited, although the majority felt that they would be prepared to use computers to vote and interact with government should computer access be more widely available. Most respondents, however, preferred to make their feelings known on an issue by writing to someone. The preferred source of government information was the public library (43.2%), with only post offices and CABx figuring significantly otherwise. For most respondents it was the accessibility of the public library and the fact that they visited it regularly that made libraries their preference. 72.5% of respondents were public library members.

The majority of respondents felt that public libraries were suitable places for finding information on their local council, consumer issues, welfare benefits, jobs and careers and housing, although there was little evidence that many had approached libraries for such information in the past. Very significant numbers were unaware that public libraries held European Union information, government publications, and information on health care, legal rights and educational opportunities, thus providing evidence of a real need for libraries to promote more effectively their collections and expertise in such areas.

References

1. Marcella, R., and Baxter, G. A national survey of the citizenship information needs of the general public. *Aslib Proceedings* 51(4), 1999, 115–121.

2. Office of Population Censuses and Surveys. *1991 Census Small Area Statistics.* Available through the University of Manchester's Manchester Information Datasets and Associated Services (MIDAS) http://cs6400.mcc.ac.uk/ (visited March 1998).

3. Office of Population Censuses and Surveys. *Standard Occupational Classification.* London: HMSO, 1990 (volumes 1 and 2), 1991 (volume 3).

4. Office for National Statistics. *Annual abstract of statistics, 1997 edition.* London: The Stationery Office, 1997.

5. Office for National Statistics. *Social trends, 1997 edition.* London: The Stationery Office, 1997.

6. Office for National Statistics. *Regional trends, 1997 edition.* London: The Stationery Office, 1997.

7. Austin, T., ed. *The Times guide to the House of Commons, May 1997.* London: Times Books, 1998.

8. Rallings, C., Temple, M., and Thrasher, M. Participation in local elections. In: Pratchett, L., and Wilson, D., eds. *Local democracy and local government.* Basingstoke: Macmillan Press, 1996.

9. Morgan, R., ed. *The Times guide to the European Parliament, June 1994.* London: Times Books, 1994.

10. Library and Information Commission. *New library: the people's network.* London: LIC, 1997.

Information retrieval techniques for evaluating search engines: a critical overview

Monica Landoni and Steven Bell
Department of Information Science, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH, UK
monica@dis.strath.ac.uk

The objective of this paper is to highlight the importance of a scientifically sounded approach to search engine evaluation. Nowadays there is a flourishing literature which describes various attempts at conducting such evaluation by following all sort of approaches, but very often only the final results are published with little, if any, information about the methodology and the procedures adopted. These various experiments have been critically investigated and catalogued according to their scientific foundation by Bell [1] in the attempt to provide a valuable framework for future studies in this area. This paper reconsiders some of Bell's ideas in the light of the crisis of classic evaluation techniques for information retrieval and tries to envisage some form of collaboration between the IR and web communities in order to design a better and more consistent platform for the evaluation of tools for interactive information retrieval.

Introduction

This paper begins by reporting on the inadequacy of classic evaluation techniques for Information Retrieval (IR) when applied to interactive IR systems. It continues by highlighting some of the alternative approaches promoted in the recent years as an answer to this crisis. In particular the chaotic situation in which evaluation of web search engines lies is clearly pointed out and a set of guidelines, mainly borrowed from interactive IR evaluation studies, is provided to assist researchers in this demanding task. This paper hopes to encourage collaboration among the IR and web communities in order to profitably exchange expertise and eventually manage to survive the chaos!

Evaluation of interactive information retrieval

Classic evaluation strategies for IR traditionally focused on the main criterion of relevance and two very popular and as such much debated measures: precision and recall. Labelled as an ineffable concept by Blair [2], the various attempts at defining relevance have resulted in metaphysical considerations, where the emphasis is often on the subjective versus objective nature of this entity and on its dependency on temporal and cognitive factors. But this has never stopped the IR community from using and very often abusing it as in the TREC experiment. With the evolution of interactive IR there is a growing number of researchers in the IR community, well represented in their efforts by the MIRA consortium, which believe this issue needs more attention since it is probably the way we could be able to understand why users do prefer one system on another and do not really care for precision and recall. Thus while the IR community is busy trying to redefine and reassess the need for criteria and measures, the web community starts a rather chaotic search for parameters to evaluate the performance of various search engines. It looks like the two communities have a lot in common and could really take advantage from some collaboration; this paper discusses how this could be made possible.

Looking for new criteria, measures and procedures

Various efforts have been made in order to attract the attention of the IR community on the ambiguity and weakness of founding an evaluation methodology completely on an 'ineffable' criterion such as relevance. Mizzaro [3] has pointed the attention of the IR community to the multi-faceted nature of relevance, while Greisdorf and Spink [4] have been reflecting on its various shades, contesting its classically binary values. Borlund and Ingwersen [5] have highlighted the subjective nature of interactive IR where user needs change according to the various situations/tasks the user is asked to perform. The role of users has been widely recognised as crucial, and the famous ASK (Anomalous State of Knowledge) which Belkin [6] described as being the real problem in IR is now an essential component of IR models. Reid [7] has gone further on this line and elaborated a task oriented methodology for evaluation of IR systems. All these efforts could contribute to solve one of the crucial problems for web users: where to find useful information and how?

The WWW problem

The Web is a huge, open, multicultural, multilingual, almost uncontrolled and ever growing repository of information. One of the most serious problems faced by web users at the moment is to be able to retrieve valuable information by fishing it out of a huge sea of neither regulated, nor guaranteed, dynamic data.

A meaningful contribution is expected from the improvement of the tools already available for searching the web (web search engines) after evaluating them appropriately. And here is where some guidance could come from the experience of the IR community. It is clear that standard IR techniques would need to be revisited in order to be applicable to such a different reality, but it is also clear that the experience accumulated since the Cranfield experience [8] could be used to bring some order into web search engine evaluation.

Evaluating search engines: guidelines

Various attempts have been made to evaluate web search engines; they have been mainly based on comparisons aimed to show the superiority of one tool on another, often guided by commercial interests. For an in depth survey on this issue see Bell [1]. The result of his research was the definition of a framework for the evaluation of search engines, which takes into consideration the complexity of web material and the user-centred new philosophy in IR evaluation. In particular a set of flexible and proceeding guidelines were identified for evaluators to use, these can be summarised in the following steps:

1. Assess the scenario: decide what is to be tested, for whom and why.
2. Review criteria and related measures: go through the list of selected criteria and measures assessing which ones and what combination will fulfil research objectives.
3. Analyse the search engines: collect information about the tools involved in the evaluation.
4. Definition of the experiment: design an experiment that takes into account the objectives and the environment assessed in the previous three steps.
5. Analyse the results: interpret the findings of the experiments in terms of objectives and expectations as defined in step one.

Assessing the scenario

This means set the research objective, according to Robertson [9] which describes five purposes behind experimentation:

- distinguishing between systems;
- finding out how to improve systems;
- testing design ideas;
- comparing with standards or success criteria;
- to find or test general principles.

So for search engines is the purpose of the test to determine which of a series of engine is the best? Is it to examine one for its usefulness? Are a number going to be examined to see how search engines can be improved? Is it to see the differences between a manually constructed index, compared to an automatically produced one? The answer to these questions will determine the character of the experiment, influencing which measures are to be concentrated on and how they are put into operation.

Reviewing criteria and related measures

Bell [1] gives an in depth and critical analysis of the various criteria and related measures used for evaluating search engines, this paper reports only the most popular and representative and the way they can be applied to search engines. Relevance is historically the criterion for evaluating IR systems and it is naturally considered a key factor in evaluating search engines even if it had to be reinterpreted carefully (see Winship [10], Hsieh-Yee [11], Leigthon and Srivastava [12]) to take into account some, or possibly all, the following factors which characterise web documents:

- duplicates may be presented;
- documents might also appear twice under a mirror site;
- links may have become inactive since being added to the database;
- pages of just links may be returned and the initial pages returned may actually lead to more relevant material;
- the pages may be written in a foreign language.

Besides few other criteria have been used together with relevance in order to keep into consideration the interactive/user-centred nature of web searches such as: usability/usefulness and satisfaction. While satisfaction can only be assessed by using direct users' feedback (e.g. questionnaires, etc.) a set of measures can be introduced in relation with relevance and usability. Here is a list of those we suggest to keep into account:

Precision: The most popular and traditional quantitative measure from IR. Metrics of search engine precision can follow three angles:

- Giving the number of relevant pages retrieved.
- Providing an indication of the total relevance of the set.
- Indicating the relevance along with the ability of the ranking algorithm.

The effects of using these different metrics is yet unclear, but given the complexity and ambiguity of making relevance judgements it would be sensible to regard the choice as a relatively non-critical issue.

Recall: The second most traditional quantitative measure related to relevance the same as precision. Clarke and Willett [13] have adapted the traditional recall measurement for use in the web environment by giving it a relative flavour. Their methodology can be utilised as follows:

- Pool all relevant documents retrieved by each service under inspection.
- Search each service for any relevant documents found in other services but not retrieved by it using a paragraph or so from the text of the document in question.
- Any documents found to be available to the service, but not retrieved, should be counted and added to the initial total of relevant documents found to form the denominator of the calculation. The numerator is the initial number of relevant documents retrieved.

Although this way of measuring recall is very rough and can be laborious it will provide a good approximate and, as with precision, produces a concrete quantitative value for comparisons.

These two measures are the main ones in their explicit analyses of the ability of a search engine to retrieve relevant documents, but there are a few other measures that help build up a full picture of the abilities of each service.

Relevance-ranking: This is still a relevance related measure but fitted for search engines. If only one page of documents is being viewed by some people then the ability of a search engine to put relevant documents in those first ten could be an important factor in its success or failure. Two measures have been reported in the literature concerning this aspect of search engine performance [11, 14]:

- Relevancy-ranking score is calculated by dividing the number of relevant documents

in the top half of the results by the total number of relevant documents in the set.
- The most-relevant-item score is calculated by giving a value to the document regarded as most relevant according to its position in the results set.

Coverage: Another relevance derived quantitative measure adapted to search engine needs. Following on from the work undertaken to calculate recall, Clarke and Willett [13] also propose a method for calculating the coverage of a search engine for any particular query. The total number of relevant documents available in a search engine's database (the denominator in the recall measure) should be divided by the total relevant documents found in all the search engines' databases.

Directory assessment: This is a qualitative measure related to usability. An alternative approach taken from the mainstream information science approach of evaluating information resources as discussed in Cooke *et al* [15]. The information gleaned from their study is qualitative and discursive; Cooke *et al* split their assessment into four sections as follows:

- Resource selection: what criteria warrant a particular site to be included in the directory?
- Level of description: what does the directory provide to alert the user to what content the referenced site has? Does the directory in question offer an analytical, critical review or indeed any description at all? Do the site references have dates of compilation and subsequent review on them?
- Subject classification and organisation: does the directory follow a systematic procedure, thus enabling efficient navigation?
- Review criteria: how sites are included in the directory, assessed either in ratings or information appearing in the reference. These criteria can easily be slanted to any particular user group by bringing out certain points more.

Noise: This is a quantitative usability linked measure. Duplicated, inactive, mirror links, any or all of these three things can be checked for, alone or altogether, depending on the researcher's point of view. They provide important information on the amount of noise that can be expected by a user, according to what steps have been taken to ensure regular updating of the search engine's database or to screen out identical documents.

Accessibility: This measure belongs to the usability area too. Users have to be confident that a site can be accessed and that a search can be conducted without interruption. What is proposed under the title of accessibility is a check to see how often the search service is unavailable or sends out an error message.

Analysis of the search engines

It is crucial in order to design effective evaluation experiments to be able to collect as much information as possible, subject obviously to availability, about the tools under evaluation, in particular it is recommendable to acquire the following:

A textual description of the service: A short paragraph describing the service, pointing out features that may make it stand out from the others. For instance users may be interested in who provides the service, e.g. in the case of AltaVista it is the Digital Corporation, how long it has been operational or any features that stand out, such as Northern Light's Custom Search Folders.

Database description: A description of the database collated by the service. It can contain the following information:

- Size of database, unfortunately having to use figures provided by the search engine producers, usually given as a number of web pages.
- Whether the database is constructed manually or automatically, i.e. if people select which sites to be included or a robotic computer program.
- Which areas of the Internet are indexed, e.g. Web, Usenet, ftp.
- If automatically constructed what penetration does the robot have, i.e. how far into any one site does it delve.

- How often the database is updated and also how often is it completely refreshed.
- What parts of documents are indexed, e.g. is it full text or partial, is link information recorded.

Input arrangement: A short detailing of what different search screens the user has to choose from, it is related to usability issues.

Search facilities: A list of exactly what techniques are available for the searcher to use in their interrogation of the database. It would be useful for the Boolean operators in use (if any) to be listed next to their traditional counterpart (AND, OR, NOT) as a quick reference. Other facilities could be listed as encountered, things like preference searching, proximity searching, truncation and field searching or limiting. An important aspect is the existence or not of a directory arrangement. Lastly what the default operator is if users were to just enter a list of keywords.

Output: A description of the information users are presented with as the results of their search. Thus a list of the components of each record provided should be made, such as URL, length of document, date etc., as should an explanation of how any abstract or summary is produced.

Ancillary services: A list of the services attached to the search engine could also be useful, including such features as entertainment 'channels', email or chat facilities.

Definition of the experiment

The design of the experiment is based on the following elements:

Queries: Researchers can develop a greater degree of reality by using queries from real users from the particular group chosen, they can have the users themselves provide the search statements and get the users to complete the searching. Using any of these strategies will improve realism used alone if full user participation is not possible. Researchers are advised to use as many queries as possible and run them as close to concurrently as practicable.

Relevance model: Criteria for relevance must be drawn prior to the assessment of returned documents. Duplicate, inactive and mirror links should all be given non-relevant rating as all can be considered to be irrelevant to users. If any links cannot be retrieved because of server errors then they should be retried on a number of subsequent occasions and if still not retrieved then given a zero rating. Any foreign language material that cannot be assessed should be replaced with another document from the returned set.

Judges: If possible use the users that produced the queries originally after being issued with guidelines. If this is not possible collect background information from the source of queries to assess usefulness as well as topic relatedness.

Documents: Only the original web documents should be used for assessment, as references retrieved by search engines are currently unreliable. Following user studies it is suggested that the top 20 documents retrieved should be used in measure calculation.

The results of the experiment will be analysed according to the criteria and measures discussed in step two and finally conclusions will be drawn.

Analysis of the results

The data obtained from the experiment should be analysed according to the purpose of the experiment. The scores for each query should at least be averaged and the means calculated for each of the search engines being used to allow for comparison but other statistical tests may be required to dig out more information. The question of statistical difference is a difficult one. It is certainly a very powerful tool for verifying results and therefore for making inferences, but it is obvious from the results of previous tests that the small number of queries being used meant that the results were inconclusive although putting search engines into a higher or lower performance band.

Conclusions

The aim of this paper is to provoke discussion on some of the open problems faced by the IR and web communities and promote collaboration between the two communities, in what the authors think are commonalties and possible areas of interaction. In particular we have just described how the experience in IR evaluation could be used profitably by the web community. At the same time the web community could provide useful advise and examples on how to deal with non-standard textual collection, something the IR community is finding hard to cope with. Finally, it is necessary to mention and emphasise that quite realistically this collaboration alone is not going to magically solve all problems but would, in the authors' opinion, represent a valuable starting point.

References

1. Bell, S. *Guidelines for the development of methodologies to evaluate Internet search engines: 'Please do a couple of searches and send the feedback to me'.* Glasgow: Department of Information Science, University of Strathclyde (MSc Thesis), 1998.

2. Blair, D.C. *Language and representation in information retrieval.* Amsterdam: Elsevier Science Publishers, 1990.

3. Mizzaro, S. How many relevances in information retrieval? *Interacting With Computers.* 10(3), 1998, 305–322.

4. Greisdorf, H. and Spink, A. Regions of relevance: approaches to measurement for enhanced precision. In: Landoni, M., ed. *Draft Proceedings of 21st Annual Colloquium on IR Research.* Glasgow, Scotland, 1999, 1–21.

5. Borlund, P. and Ingwersen, P. The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53(3), 1997, 225–250.

6. Belkin, N.J. Interaction with texts: information retrieval as information-seeking behaviour. In: *Information retrieval '93.* Von der Modellierung zur Anwendung. Konstanz: Universitaetsverlag Konstanz, 1993, 55–66.

7. Reid, J. A task-oriented, non-interactive evaluation methodology for information retrieval systems. *Journal of Information Retrieval*, 2(1), 2000, 113–127.

8. Cleverdon, C.W., Mills, L., and Keen, M. *Factors determining the performance of indexing systems.* Cranfield Project, Cranfield: Aslib, 1966.

9. Robertson, S.E. The methodology of information retrieval experiment. In: Sparke-Jones, K., ed. *Information retrieval experiment.* London: Butterworths, 1981.

10. Winship, I.R. World Wide Web searching tools: an evaluation. *Vine*, 99, 1995, 49–55.

11. Hsieh-Yee, I. The retrieval power of selected search engines: how well do they address general reference questions and subject questions? *The Reference Librarian*, 60, 1998, 27–47.

12. Leighton, H.V. and Srivastava, J. *Precision among WWW search services (search engines): AltaVista, Excite, Hotbot, Infoseek and Lycos.* 1997. http://www.vinona.msus.edu/library/webind2/wi2pt1.html (visited 29 August 1997).

13. Clarke, S.L. and Willett, P. Estimating the recall performance of web search engines. *ASLIB Proceedings*, 49(7), 1997, 184–189.

14. Ding, W. and Marchionini, G. A comparative study of web search performance in global complexity: information, chaos and control. In: *Proceedings of the 59th Annual Meeting of the American Society for Information Science.* Baltimore, Maryland, 21–24 October, 1996, 136–142.

15. Cooke, A., McNab, A. and Anagnostelis, B. The good, the bad and the ugly: Internet review sites. *Online Information 1996: the proceedings of the 20th International Online Information Meeting.* London, 3-5 December 1996, 33–40.

CONTENTS

133–137 'The liveliest of corpses': trends and challenges for the future in the book publishing industry
Iain Stevenson

134–142 Quality standards in drug and medical information departments
Sadia Khan and David Bawden

143–149 Implementing the 'New Library: The People's Network' and the management of change
Alan Boughey

150–160 Facilitating scientific communication for industrial innovation
Paul M. Evans

'The liveliest of corpses': trends and challenges for the future in the book publishing industry

Iain Stevenson
Director, Publishing Studies, Department of Journalism, City University, Northampton Square, London EC1V 0HB
W.I.Stevenson@city.ac.uk

Reviews the perceived and actual threats to the book publishing industry arising from electronic storage, management and distribution of information. It argues that publishing has always been primarily a content-management industry and if not entirely free from technological obsolescence, it has at least been relatively insulated from its worst effects by concentrating on the maintenance of quality and authority of the information it provides. Greater challenges may arise from other quarters, Including globalisation, threats to intellectual property regimes, and changes in retailing and distribution, and publishers must think about the impacts of these on their businesses.

In September 1998 the following announcement appeared in a San Francisco newspaper:

> *Gutenberg Technologies Inc* announce the launch of a revolutionary new platform for information delivery for the consumer, business, scientific and educational markets. It features an intuitive user-friendly interface that an untrained user can master in moments. Random access and multiple-entry protocols to the stored information allow copying, transfer and storage. Search and retrieve functions are provided on opening. Memory life is guaranteed at 300 years. The interface features write-on technology that allows the user to highlight or add additional information. Text and graphics can be interdigitated without changing programs. The platform runs on existing commonly-available software, carries its own power source, is environmentally-friendly to EPA standards, weighs only 16 ounces and is Y2K compliant. This revolutionary new device will be called BOOK and is available from your neighborhood bookstore.

This was, of course, a clever advertisement for a bookshop but draws attention to the extraordinarily robust and flexible technology that the print-on-paper codex represents. Ever since the Flemish itinerant craftsman William Caxton and his wonderfully named assistant Wynkyn de Worde set up his heavy wood and metal German designed printing press in the shadow of St Paul's Cathedral in 1476, British book publishing has been a major cultural industry that has dominated world learning, leisure, information, business and commerce for over half a millennium. Like many familiar technologies, we fail to appreciate just how sophisticated, flexible, many layered and multi-purpose this combination of cotton fibre, lignin, fabric, mineral inks, glue, thread and plastic that make up the modern book can be.

Yet, it is generally considered that as at no other time in its history the book and by implication the publishing industry is under threat of immediate and complete extinction. New technology will send the denizens of Grubb Street to the same fate as tallow boilers, hansom-cab drivers, crossing-sweepers and telegram boys. Of course books have been doomed by new technologies before, by films, radio and television. In reality the book has survived each of these threats and these media curiously have probably done more to enhance and widen book use than damage it. The beneficial effects of a blockbuster new film or a major new TV series on book sales is demonstrated daily in bookshops. One publishing company (the aptly-named Boxtree) concentrates its business on 'TV tie-ins' and makes many millions of pounds of sales on books that reflect television programming. However, the recent damage to Dorling Kindersley's business by their over-estimating the demand for 'Star Wars' related publishing, which may have cost them £25 million in losses, may be salutary in establishing the limits to screen-related exploitation.

But surely it is different now? The Internet, and in particular its graphic component the World Wide Web, allows almost instant access (in this hemisphere at least until lunchtime when California wakes up) to millions of pages of information, entertainment and education. As an aside, and on a point that will be returned to, it is interesting to note that we still think intuitively of a screen display as an analogue of an old-fashioned print-on-paper interface. Moreover, Internet pages all come with interactivity (of a kind) and are all free of charge (for the time being). And waiting in the wings are new devices, even faster, more responsive and more mobile. WAP (wireless application protocol) technology, which will very shortly allow Internet access via mobile phone networks, will make deskbound computers redundant and will at a stroke solve the portability problem that still give books their slim superiority over electronic information retrieval and display systems.

And yet...... Rather eerily just six months after the advertisement above was published, Rocket eBook [1] was launched. Essentially a palm held computer, Rocket eBook is causing a great deal of excitement in the information and book retailing worlds. With a large backlit screen, easy to read in all lights and a weight of 22 ounces, this device can be loaded with the text and illustrations from up to ten full length books (purchased over the Internet from the book retailer Barnes and Noble). These can as its website proudly announces '...[be] read...just as you would a paper book'. Its further advantages are that you can take it 'anywhere you go to work, on trips, to the beach, or to bed. ...[You can] make margin notes, underline special passages, and bookmark pages.' Most wondrous of all, you can even '[b]egin reading instantly.'

And all this is available for only $199.00 plus shipping, plus the cost of the downloaded books, and, of course, access to a compatible power source.

A semiologist might be tempted to conclude that the claims for Rocket eBook have something of the insecure behind them: that the new proud all-conquering technology it represents only feels that it can establish its legitimacy by aping the obsolescent despised low-tech of print-on-paper. The American critic Sven Birkets in his thought-provoking essays on reading in the Internet age, *The Gutenberg Elegies* [2], identifies what he calls 'the shadow life of reading'. He points out that the reader's engagement with the printed page is a much more subtle and active relationship than the more passive information retrieval gained from computer, television or motion picture screen. The act of reading from print-on-paper is an aesthetic almost mystical experience in which the author and reader commune, with the latter bringing almost as much to the experience as the former. The emotional, intellectual, perhaps even spiritual engagement with the printed page that we have all experienced is, Birkets argues, unique to the medium. We can all report having become 'lost in a book', being taken to some other place by the power of imagination catalysed by the magic chemistry of typography. Perhaps this is why book-related terms and metaphors are so deeply instinctively embedded in our consciousness that they remain essential to our understanding of electronic media. Perhaps eBook and its developments can offer us the same engagement with print and the same transportation to other worlds. Perhaps not.

But even so, surely books and the publishing industry are finished this time, new technology will blow the old-fashioned, hidebound (again another book-derived metaphor!) traditional book business to smithereens and print-on-paper will become museum curiosities and libraries will be reduced to pocket-sized mobile phones. The perspicacious Finns have surely made the correct economic forecast for the future and are busily engaged in replacing their paper-pulp industry (and thus preserving their boundless forests) with world leadership in the mobile phone business.

Yet for a terminally ill industry, publishing is a remarkably lusty invalid. You might even think it is perversely healthy. Between 1990 and 1997, publishers' sales in the UK alone grew from £1364 million to just over £2000 million [3]. Add to this exports of £844 million (up from £531 in 1990) and you have a dynamic growing successful industry with an employment of over 90,000 people and contributing a sizeable chunk of GDP. (Which totaled just over £600 billion in 1997.) It is interesting to note for comparison that publishing is almost a quarter of the size of the entire agricultural sector. If publishing is under threat of imminent collapse, perhaps we should organise a literary alliance and march on Parliament demanding subsidies on blue stockings and tweed jackets with leather patches.

The truth is, of course, that predicting the death of publishing in the face of the Internet or other e-technology is misplaced and arises from a fundamental misunderstanding of what publishing actually is and does.

Publishing is not, and never has been in its history, primarily about technology but is about information management. What publishers do is create content, manage it, add value to it, stretch it, recycle it, translate it, and transmit it. In truth, publishers have been successful from a very early stage in distancing themselves from delivery technology, and leaving the economically questionable job of managing ever-obsolescing technology to others. Publishers generally ceased to be their own booksellers in the 18th century, their own printers in the 19th century, and in the 20th century, they have coped with a remarkable technological shift in printing and book manufacturing technology from hot metal to cold setting. As we enter the 21st century the impact of new technology will only make it more essential than ever for publishers to concentrate on their content management skills and leave the mechanics of distribution and delivery to others. In a telling phrase, they should seek to be 'the sewage not the sewer'.

This is not to say that publishing does not have some important challenges to face, not least technological change and development, and this article will conclude by briefly considering some of these.

Globalisation

The first and most important is globalisation. Like all information businesses, and indeed most industries, publishing has undergone a rapid and dramatic global expansion of the arena of its market and in its ownership. A glance at the family tree of British publishers will show that the original complex network of mainly private and independent companies have been increasingly incorporated into major multi-national corporations. Of the major players, only one (Pearson) is British owned, and its culture is largely American or at least trans-Atlantic. Of the others, two are German (Holtzbrinck, Bertellsman), one Canadian (Thompson), two Dutch (Elsevier, Wolters Kluwer) and two American (HarperCollins, Little Brown). Indeed the two largest 'truly British' publishing companies are institutional and presumably impervious from takeover (Oxford University Press, Cambridge University Press).

Although ownership is increasingly global, globalisation is not a new thing in publishing. British publishers, much more than Americans, French or Germans, have always seen their markets as worldwide: Macmillan and Longman were publishing for and in India, Africa and Asia from the late 19th century and a specifically global educational market that Britain dominates is English Language Teaching for non-native speakers (referred to in the jargon as ELT or ESL), worth £260 million in 1997, or almost a third of book exports. A major factor in globalisation is the dominance of English as the medium of communication throughout the world.

Globalisation will mean that traditional territoriality will break down but that may provide opportunities as well as threats.

Intellectual property

The second major challenge is in the field of intellectual property. The basis of all content management is a clear system of ownership and reward for use. Despite the Web, information never has been, never could be, and never should be free. The value of information depends both on its intrinsic nature and its timeliness. It has to be both authoritative (accurate, true) and accessible. What publishers do essentially is guarantee the former and ensure the latter. This of course costs, and to make sure it is done well, publishers have to be paid. Distribution technologies are changing and it is likely that much published information will be web-distributed in future. As information is merged in databases, who owns and guarantees which bit becomes of vital importance.

Technological change

This introduces the third challenge, which is, of course, technological change. Publishers have always dealt with this and quite effectively and many new technologies favour publishers. 'Watermarking' technologies for identifying text-ownership (of which the Digital Object Identifier is an example) will allow flexible, trackable dissemination of published works in many forms and media, and will certainly widen access and use. Print-on-demand digital printing that makes print runs as low as five economic and means that no book need ever go out-of-print again. It will be possible using this technology for any book to be individually custom-printed and bound for a customer at a bookshop terminal while they wait.

Book distribution and retailing

And finally, this introduces the dramatic transformation that is overtaking book distribution and retailing. Perhaps the most rapid and dramatic change to affect publishing has been the change in retailing and distributing the product. The creation of enormous 'book palaces' like the Borders Bookshop in Glasgow and Waterstones in Simpsons in Piccadilly, based on United States models, have completely revolutionised the image of bookselling. Interestingly Birkets, referred to above, worked in the ur-Borders in Ann Arbor, Michigan, and is very interesting on the origin of the merchandising concepts pioneered by the brothers Border.

Bookstores are now fashionable, young, exciting places, with coffee bars and 'encounter zones'. They are the places to be seen, to meet one's friends, pick up new ones and perhaps occasionally buy a book. The famous Indigo store in Toronto combines political-correctness with a fashionable meeting place. It champions Canadian culture and writing in the face of American invaders and, also, is the favoured meeting place of the Ontario capital's young, attractive and intellectual. So much so that it is said that prospective browsers dress in their best (or their sexiest) to visit its shelves.

Perhaps the most dramatic change in retailing has been the rise of Internet bookstores. The most well known and glamorous of these is Amazon.com (with its British Amazon.co.uk and German Amazon.de subsidiaries) which offers immediate access to over three million titles, impressive 'shopping cart' and '1-click' payment technology and a user friendly interface. The service is certainly impressive: the author ordered a book recently from Amazon.co.uk on one Wednesday evening at 11.30 pm which arrived by mail first post on Friday morning. A report recently published predicts that 17 per cent of book purchases will be made online by 2004 – although only one-sixth of this would be 'new' business, the rest being captured from traditional booksellers.

Conclusion

Recent corporate re-alignments, like the American Publishing Group Time-Warner forming partnerships with Internet Service Provider America On-Line (AOL), and with Music Industry giants EMI thereby apparently attempting to create a seamless continuum from content creator to distribution channel to user, might be thought to undermine the author's arguments about publishers surviving better independent of e-distribution. However, I would submit that it in fact underscores my contention. The Time-Warner move is primarily about access to film, television and particularly performance music which is very different in culture, audience relationship and patterns of use than literature, even broadly defined. The print environment will certainly change in the face of the electronic challenge, but a confident prediction is that the heirs of Caxton and Gutenberg can look forward to as dynamic, lively and interesting half-millennium as that from which they have just emerged. After all, you are reading this article as print-on-paper, aren't you? *Aren't you?*

References

1. The Rocket eBook. Available at: http:/www.rocketbook/com/enter.html

2. Birkets, S. *The Gutenberg Elegies: the fate of reading in an electronic age*. London: Faber and Faber, 1994, 229.

3. *Book Trade Yearbook 1999.* Publishers' Association, 1999. (This and other statistics quoted in this article are from the *Book Trade Yearbook 1999.*)

Quality standards in drug and medical information departments

Sadia Khan
Royal Pharmaceutical Society, 1 Lambeth High St, London SE1 7JN
skhan@rpsgb.org.uk

David Bawden
Department of Information Science, City University, Northampton Square, London EC1V OHB
dbawden@soi.city.ac.uk

Quality standards in the enquiry answering function of health service drug information units and of pharmaceutical industry medical information departments are studied by an interview/questionnaire survey. Factors examined include the use of standardised and documented procedures, adherence to published guidelines, methods of evaluating quality, and the impact of ISO 9000 certification.

Introduction

The provision of information on medicines is a vital part of modern healthcare [1]. Two main sectors are involved. Within the health service, drug information centres, based in pharmacies of regional or local hospitals, provide information on medicines to healthcare staff in their area, while pharmaceutical companies maintain medical information services, providing information on the products of that company, to healthcare workers, and others, nationally or, in some circumstances, worldwide [2, 3, 4]. Although the environment and raison d'etre of the two kinds of unit are rather different, an overview of the literature concerning the enquiry answering activities of the National Health Service (NHS) and industry drug/medical information units has suggested that, broadly speaking, the types of enquiries dealt with by the two sectors are very similar [5].

Quality assurance, and associated methods of performance measurement, have come to be seen as of great importance for information services of all kinds [6]. In particular, certification for the ISO 9000 quality standard has been seen by some as the strongest demonstration of service quality [7]. In the case of specifically medical and pharmaceutical information, quality is of particular significance; there is a clear analogy with the trend towards evidence-based practice in all aspects of health-care (see, for example, Muir Gray [8], Pyne [9], Scherrer and Dursch [10]). A methodology for applying the ISO 9000 standard to the task of searching the medical literature has been presented [11].

For the specific issue of provision of medicines information in Britain, there are two main relevant sets of quality procedures: those included in the *Drug Information Manual* of the UK Hospital Pharmacists Drug Information Group (DIPG), and those stated in the *Guidelines on Standards for Medical Information Departments* published by the Association of Information Officers in the Pharmaceutical Industry (AIOPI). The first covers the provision of medicines information within the health service, and the second covers the activities of medical information departments within pharmaceutical companies. Both were initially created in the early 1990s, and have been revised with experience.

The DIPG guidelines are very detailed, and deal with criteria, parameters and standards for accessibility, resources, usage of enquiry services, analysis of workload statistics, customer satisfaction surveys, and peer assessment of the quality of randomly sampled queries. The AIOPI guidelines are necessarily less detailed and more generic than the other two, and are presented as advice on good practice within a diverse sector. They cover accessibility and availability, timeliness, information resources, qualifications and training of staff, and a variety of audit techniques and performance indicators.

The purpose of the study reported here was to assess the current situation with respect to quality assessment of information provision, specifically enquiry answering services, in drug and medical information services.

The survey

The survey was carried out in two phases. First, a semi-structured interview was carried out with information professionals in four pharmaceutical information services, chosen so as to reflect the diversity of information provision in the area. The results obtained from these interviews were combined with information obtained from other sources – literature searching, examination of published quality manuals etc. – to draw up a series of detailed case studies of variant practice.

In order to assess the generality of the results, a postal questionnaire was constructed, and mailed to 50 hospital pharmacy drug information centres, and 50 medical information departments in pharmaceutical companies.

Interviews

Four information services were chosen for interview:

- a regional drug information centre, having ISO 9000 certification;
- a regional drug information centre, known as a well-established and efficient centre, but not having ISO 9000 certification;
- a large and well-established pharmaceutical company medical information service, having ISO 9000 certification;
- a small and newly established pharmaceutical company medical information service, not having ISO 9000 certification.

Investigation of this diverse set of centres was likely to draw attention to a wide variety of issues and attitudes.

Interviews were carried out by telephone, and lasted about 40 minutes. The outline interview format is shown in Appendix 1.

Questionnaire

Fifty hospital pharmacy drug information services – including both regional and local centres, and 50 pharmaceutical company medical information departments – including companies of all sizes, were chosen randomly from the directories of the UK Drug Information Pharmacists Group and the Association of Information Officers in the Pharmaceutical Industry respectively. To each of these a written questionnaire was sent; the outline of the questionnaire is shown in Appendix 2.

The questionnaire was sent by mail, if a response was not received within three weeks, a reminder letter was sent. 45 out of 50 services responded for both health service and pharmaceutical industry, a response rate of 90%.

Results

Interviews

The pictures built up in four case-studies based on the interviews revealed a pattern of practice which, while diverse, was perhaps more consistent than might have been expected from four units selected for their 'formal' diversity.

All four units took the quality of the enquiry answering service seriously, and all had clear procedures in place; all were based on published guidelines (DIPG and/or AIOPI), and for all, except the small industrial unit, these procedures were documented. All made some attempt at evaluating the adequacy and appropriateness of these procedures and a range of methods were used. The major distinction was in the extent of quality assessment, in the number of different methods used, and in the depth of analysis employed. This distinction appeared to be largely resulting from the availability of resources, rather than from any difference in philosophy; in particular, the presence or absence of ISO 9000 certification did not seem to make much of a difference.

Questionnaire

The pharmaceutical industry and regional health service units had an average of four professional staff, while the equivalent figure for local health service units was two. The average number of enquiries handled per month varied considerably: for the industry it was 651 (with a range of 30 to 2250), for NHS regions 317 (range 100–400) and for the local NHS 115 (range 25–300).

All the NHS units had procedures for query answering, as did 45 out of 50 industry units. The percentages of those for which the procedures were stated to be documented were 100% for the health service and 66% for the industry. All of the NHS units followed the DIPG guidelines, as did 22% of the industry units; 93% of the industry units followed the AIOPI guidelines. (This overlap indicates the probable similarity of work carried out in the two sectors.) Thirty per cent of the industry units, and 10% of NHS, also followed local guidelines and procedures.

Ninety-one per cent of the NHS units and 87% of the industry assessed the quality of enquiry answering. The methods used were reasonably consistent across the two sectors; there was an indication that regional NHS centres made use of a wider variety of methods than local centres, presumably due to their higher levels of staff and resources generally. The most popular method, with more than 80% in both cases, was analysis of workload statistics; its popularity is unsurprising, given that this information is essential for a unit manager in justifying activities. User feedback questionnaires, evaluation of sources used and evaluation of quality of answers were used by more than 60% of the NHS units, but less than 45% in the industry; this may well be because the DIPG guidelines provide convenient templates and standard forms for these methods. Conversely, peer review of answers by individual experts and agreed timescales for answers were used by more than 70% in the industry, considerably less in the NHS. The other methods mentioned in the questionnaire – technology for monitoring telephone rings, telephone surveys of user satisfaction, committee peer review of answers, and 'other methods' – were used by less than 20% in both cases. The other methods included logs of 'spontaneous feedback' such as letters of thanks, analysis of customer complaints, coaching and mentoring of staff, and enquiry logging databases.

Thirteen per cent of NHS units and 40% of industry units were in organisations where some aspect of operations was covered by ISO 9000 certification. Where this was the case, medical information procedures were included for four out of six of the NHS units, but only for one out of eighteen in the industry. In almost all cases, the reasons for seeking certification related to wider issues within the organisation, rather than the specifics of pharmaceutical information provision.

For the few units with certification, a variety of locally produced documents (operating procedures, reporting procedures, enquiry logs etc.) and also published sources (British National Formulary, ABPI Data Sheet Compendium etc.) were treated as controlled documents.

Opinions on the effect of certification were somewhat varied, though with only five units included the results are hardly significant:

	Yes	No	Don't Know
Affected workload	4	1	0
Affected resources	2	3	0
Affected user perception	1	1	2

Benefits were generally stated as improved documentation, leading to improved consistency of service, and a belief (as yet unsubstantiated) that this would lead to a better service; the only stated drawback was the increased paperwork.

Few respondents saw any likelihood of future quality innovations; where they were likely, it would be as part of a wider organisational initiative, rather than anything specific to, or led by, the medical information department.

Conclusions

The results of this study suggest that quality is taken very seriously in the enquiry answering functions of drug/medical information services. Virtually all the units surveyed had documented procedures, and observed published guidelines and procedures. A general similarity between the two sectors – health service and industry – was seen, with differences being those of emphasis and convenience, rather than principle, reflecting the commonality of the knowledge base being accessed, and the kind of information requested.

A range of quality evaluation measures was found to be used, in a reasonably consistent manner across the two sectors; the differences reflect the varying emphasis in the guidelines for the two environments. It is surprising that the coaching role – seeing how enquiries are actually dealt with – was mentioned by only one out of ninety respondents. This method might be expected to be useful in training new members of staff, as well as in an evaluative role.

ISO 9000 appears to have little impact on this sector. The small number of services involved with certified procedures means that no clear pattern of its effect could be established. It seems clear that information services are, in this respect, being led by general organisation policy, rather than taking a lead in seeking, or for that matter avoiding, certification. It is notable that one industry respondent argued that attaining certification would be duplication, since clear responsibilities, processes and procedures were already in place. ISO 9000 is clearly not universally viewed as the sole measure of quality, and the results of this survey suggest that its effect on quality may be one of perception and image rather than anything more substantive.

References

1. Robson, A., Bawden, D. and Judd, A., eds. *Pharmaceutical Information Management*. Edinburgh: Churchill Livingstone, forthcoming 2000.

2. Hopkins, F., Galligher, C. and Levine, A. Medical affairs and drug information practices within the pharmaceutical industry: results of a benchmark survey. *Drug Information Journal*, 33, 1999, 69–85.

3. Robson, A. *et al.* Medical information services in the pharmaceutical industry: a survey of pharmacists' views with recommendations. *Pharmaceutical Journal*, 256, 1996, 864–866.

4. Huntingford, A.L. *et al.* A survey of industry medical information departments' activities. *Pharmaceutical Journal*, 245, 1990, 238–239.

5. Norton, R. *Pharmacists' information needs.* Unpublished MSc dissertation. City University London, 1996.

6. Brophy, P. and Coulling, K. *Quality management for information and library managers.* London: Aslib, 1996.

7. Ellis, D. and Norton, B. *Implementing BS EN ISO 9000 in libraries*. London: Aslib, 1996.

8. Muir Gray, J.A., ed. *Evidence-based healthcare*. Edinburgh: Churchill Livingstone, 1996.

9. Pyne, T. *et al.* Meeting the information needs of clinicians for the provision of evidence-based healthcare. *Health Libraries Review*, 16, 1999, 3–14.

10. Scherrer, C.S. and Dursch, J.L. The evolving role of the librarian in evidence-based medicine. *Bulletin of the Medical Library Association*, 87, 1999, 322–328.

11. Cullen, R. and Mason, D. Quality assurance in health sciences literature searching: applying the ISO 9000 quality standard. *Health Libraries Review*, 12, 1995, 173–189.

Appendix 1

Semi-structured interview format

- How many enquiries are dealt with by your centre each month?
- How many professional staff deal with enquiries in your centre?
- Describe the quality measures and procedures used in enquiry answering.
  [Examples were given as necessary as prompts.]
- Are your quality measures based on published guidelines?

For services with ISO 9000 certification:

- What factors influenced your decision to gain ISO 9000 certification?
- How has gaining certification affected your department?
  [In terms of workload, resources, status etc.]
- Describe the materials classified as 'controlled documents' with the certification.

For services without ISO 9000 certification:
- What is your department's attitude to certification?

Describe any other quality initiatives your department is undertaking or considering.

Appendix 2

Questionnaire

Questions were:
- Number of medical staff dealing with medical information enquiries?
- Average number of enquiries handled per month?
- Are there procedures for enquiry answering?
  If YES:
  - Are procedures for enquiry answering documented?
  - Do the procedures reflect: UKDIPG guidelines; AIOPI guidelines; other guidelines (give details)?

- Is the quality of enquiry answering assessed?
  If YES, indicate which of the following are used:
  - Monitoring of telephone answering time.
  - Agreed timescales for enquiry turnaround.
  - Monitoring of user satisfaction by questionnaire.
  - Monitoring of user satisfaction by telephone survey.
  - Evaluation of documentation of sources used.
  - Evaluation of quality of answers.
  - Monitoring of workload by query statistics.
  - Committee peer review of query answers.
  - External peer review of query answers.
  - Other (give details).

- Does any part of the organisation have ISO 9000 certification?
  If YES:
  - Are medical information procedures included within the certification?
    If YES:
    - What factors influenced the decision to attain certification?
    - What medical information materials are designated as 'controlled documents'?
    - Has ISO 9000 affected medical information workload?
    - Has ISO 9000 affected medical information resources?
    - Has ISO 9000 affected users perception of medical information?
    - Describe the benefits and drawbacks of certification.

- Describe any further developments in quality management within your organisation.

Implementing the 'New Library: The People's Network' and the management of change

'Oh, what on organisational web we weave when first we start to electronically retrieve.'

Alan Boughey
Bury Reference & Information Services, Central Library, Manchester Road, Bury, Lancashire, BL9 0DG
a.boughey@bury.gov.uk

The New Library: The People's Network and other recent government initiatives have set out a new direction for public library services in Britain. The aim is to create 'hybrid' libraries, combining existing book lending and information services with extensive ICT access. With the new technology will come enhanced roles in information, learning support and reader development. Management of change theory emphasises the effects of introducing change in one area on the whole of the organisation. If this theory is applied to public libraries, significant 'knock on' effects can be expected. Some examples are taken from the practical experience of Bury Libraries.

Introduction

If, as the Chinese proverb says, it is a curse to live in interesting times, then Britain's public libraries are cursed even more than usually at the moment. During the last few years a new vision has been set out for the service, one based on existing strengths but giving much greater emphasis on providing all sections of the community with access to the new Information and Communications Technology (ICT).

Picking up on research, commissioned by the previous administration, the Labour Government has put its active support behind the ideas contained in the *New Library: The People's Network* report. A series of government initiatives are clearly pushing public library authorities in the way the Government wants them to go. The response of the organisations of the library profession has been enthusiastic, but what sort of obstacles will the New Library face over the years as it is more fully realised?

The article draws upon theoretical reading undertaken for a postgraduate Diploma in Management focusing on the subject of the 'Management of Change' combined with practical experience of involvement in a variety of ICT initiatives in a small Metropolitan Borough library service, Bury. These have included the introduction of Internet services at almost all service points, managing the establishment of an Internet site, contributing towards policy development within the service, staff training and day-to-day management of a busy information enquiry service using the Internet and CD-Roms alongside books and local history materials.

New Library: The People's Network

The *New Library: The People's Network* was produced by a Working Group of the Library & Information Commission (recently replaced by the Museums, Libraries and Archives Council) in 1997 [1]. Following a favourable response from the new New Labour Government [2], a follow-up report, *Building the New Library Network* [3] appeared in 1998, setting out the means by which some of the aims of the report could be practically achieved.

The vision set out for public libraries in these documents is one of libraries connected to a powerful network (or set of networks), offering ICT services in an integrated way with other, more traditional services, such as the lending of books and other materials, information enquiry services and support for learners. As now, all this will be delivered in a local, community based way and the report goes out of its way to dampen fears that books will be dropped. The intention is not revolutionary: public libraries will be doing much as they have always done, but in new ways. This is sometimes referred to as the 'hybrid library,' one which employs both books and ICT.

Three areas of work were set-out: creating content, training staff, and the network infrastructure. By the time of the second report, it was probable that content and training were both to be funded by the New Opportunities Fund (NOF), so the report was able to outline the programme's aims and the practical measures to achieve their targets in these areas. Funding was the big problem for the network infrastructure, and became the main cause of concern in the report.

Other government programmes tie into this plan, in addition to the NOF schemes for training and content creation. The National Grid for Learning and University for Industry aim to make use of this network, providing both enriched content on the Web, especially open learning materials, and a network of access points for school-age children and adults wishing to re-skill (with Community Access to Lifelong Learning centres being another new scheme to do this). Annual Library Plans are now required from each authority, and they are specifically marked on their plans to achieve the goals of the New Library. The first three years of the DCMS/Wolfson Challenge Fund provided competitive opportunities to bid for network infrastructure funding (though the fourth year has switched emphasis to reader development). The Government's push for greater accessibility to both central and local government, and the ambitious aim to facilitate electronic interaction between citizen and state (filling in forms online, etc.) imply widespread accessibility to ICT resources. Public libraries are obvious candidates to play a major role here.

What will the New Library do?

So what does all this actually mean for public libraries? Looking at the New Library report in a little more detail, the 'principal strands of content and services which will pervade the networked public library' are quite familiar:

- Education and lifelong learning (from story times which help the very young to read, to providing local access points in conjunction with higher education institutions).
- Citizens' information and involvement in society (providing access to government information and helping to achieve the Government's ambitious targets for electronic interaction between citizens and the various agencies of the state).
- Business and the economy, training and employment (providing enhanced business information for companies and individuals changing employment or dealing with consumer issues).
- Community history and community identity (making local history materials available as a part of the content creation element of the New Opportunities Funding).

Most of the above involve no major new services, doing what is done now, but using computers. The only real exception is the publishing of more local history and community information on the Web, though even this is done in paper form to a greater or lesser extent in many authorities.

The New Opportunities Fund's ICT training documents slightly alter these priorities, perhaps due to the effect of the National Year of Reading and the emphasis on basic skills – especially literacy – in the education field. The three areas for the employment use of ICT in the library environment are given as reader development, finding information and assisting learning.

Content creation

Using £50 million of New Opportunities Funding, libraries have been encouraged to become sources of digital content. The aim is to produce digital information on local history, culture, literature and other materials of use for lifelong learning in the broadest sense of the phrase. Materials specifically aimed at the curriculum are discouraged under this funding stream. Other sources of content include the University for Industry (marketed under the brand, Learn Direct), geared to online delivery of courses.

Training the New Librarian

Significant effort is to be put into training individual library staff to make sure they are up to the challenges of their new role. Even staff currently working in the information/learning support environment will need to significantly enhance their ICT skills. Those working in the 'core service' lending libraries, especially in smaller branch libraries, will generally have a harder job, not only learning ICT skills, but also the information/learning support skills often previously left to more specialised staff.

Over the last few years, most public library staff have been trained on a 'need to know' basis to use specific CD-Rom, Internet and other related products. Although many staff have taken to this enthusiastically, especially if they have a PC at home, it has left many behind – some complaining of a 'lack of confidence'. This has combined with a poor appreciation of just what levels of skill are required to an over-reaction, with staff feeling they are expected to become 'experts', a role they are clearly unable to fulfil, thus reinforcing a downward cycle of lack of confidence breeding lack of skills.

With the aim of providing wide-scale basic training many library authorities are adopting the 'off the shelf' European Driving Licence standard to address this fundamental training need. Twenty million pounds of New Opportunities Fund money will be allocated between April 2000 and March 2003 primarily to fund this programme, but library authorities will be expected to fund cover for the staff to undertake the course. But will this do the job? It is already clear that the answer will, at best, be 'partly'. A recent British Library Research & Innovation Centre Report, *Staff in the New Library: skill needs and learning choices* [4], clearly identifies other non-ICT skills required, those of the more traditional professional librarian: information enquiry handling and learning support. Just because a library assistant knows how to use an Excel spreadsheet this does not mean that they are equipped to find accurate market research data, from a variety of paper and ICT resources, for the entrepreneurial Patel family featured in the *New Library* report. This is born out by a detailed look at the 'expected outcomes' of the NOF funding, which include much on how to use ICT skills in the library.

The Internet in Bury libraries

Bury has benefited from an authority-wide connection to the Internet and, as a geographically small borough, most service points were already connected (albeit with relatively low capacity 64 bit lines). Basic PC, as opposed to library circulation system, experience had been gained from early use of CD-Roms. Provision of public access in all locations, therefore, rolled out fairly quickly starting in summer 1997. Staff use machines were also provided for enquiry use, staff training and email.

A key priority for the local authority was to ensure social inclusion. Charges, therefore, were kept low (£2 per hour) with half price access for Leisure Cardholders (a concessionary scheme). Despite the fact that many authorities are gradually moving to free access, the income generated is hard to ignore or refuse.

A 'Staff Forum' of representatives from most service points and parts of the library structure helped to 'spread the word'. Most staff viewed these developments positively as a way of broadening the appeal of their libraries (the threat of closures was very real at this point, as three part-time branches were going through a long drawn out process of closure). Although no 'technical' staff were directly employed, the council's IT service provided good support. Problems included the lack of time to undertake more than very basic training and poor security for closed buildings leading to equipment thefts.

Additional pressure on the busiest service points has been felt, as these tend also to be busiest for the Internet. Information enquiry work is one area that has suffered as an unavoidable consequence. Table 1 shows the additional workload for the Reference &

Table 1. *Workload comparisons for Reference and Information Services Section.*

Major categories of enquiry	**1994 (pre-Internet)**	**1998 (post-Internet)**
Internet	–	141
General	358	324
Business	18	40
Local and family history	102	75
Current local information	99	53
Statistics	8	9
Community information leaflets	5	5
Total	**590**	**655**

Table 2. *Use of Internet by adults.*

1998/99	**Internet hours booked**	**Issues**
Bury	1,889	353,463
Prestwich	986*	247,594
Radcliffe	495	169,224
Ramsbottom	411*	142,100
Unsworth	144*	79,754
Tottington	71*	71,541
Whitefield	N/A	95,445

* Installed during year.

Information Services (RIS) section of Bury Central Library.

Adult Internet use figures for all locations are given above (Table 2), compared with total issue figures (all categories).

Some use has been like that described in the *New Library* report. The Web has been a useful source of information for the full range of enquiries, from business to homework. Learning, at least in ICT, has been widespread, assisted by our free one hour 'Get You Started' Internet courses (500 hours of courses were provided in 1997–1998), followed-up by longer evening classes held in libraries but run by the Lifelong Learning (adult education) service. Like most other public libraries, however, a large proportion of use has been for recreational email purposes. Although this purpose has brought considerable additional usage to the service, it has created some problems including demand for very short booking periods and users with absolutely no ICT skills expecting staff to do the emailing for them.

A survey of users in the RIS section of Bury Central Library gave the following data:

What do you use the service for?	
Finding information	41
Email	16
Chat/discussion groups	8

Have you been on one of our Get You Started courses?	
Yes	7 (16%)
No	41 (84%)

How would you rate the Internet service?	
Excellent	16 (36%)
Good	18 (40%)
OK	11 (24%)
Poor	0

Staff roles have varied from the mundane (constant fixing of broken printers) to educator (informally, and as tutors for the Get You Started courses) and information 'broker' advising on the best web sites. To assist the latter, a *Guide to the Web* available in printed and online versions provides locally popular jumping off points (mainly gateway/ portal sites).

Children are provided with free access, as are the visually and hearing impaired, through the specialist units for those users. The visually impaired, in particular, benefit from screen enlargement and speech synthesis.

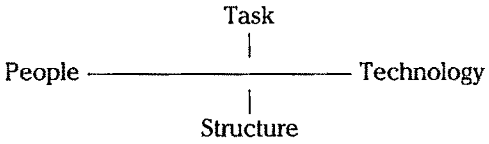
Content creation is less well advanced, but a beginning has been made. Web pages for Cultural Services, including libraries, were launched in 1998 as a part of the corporate site (http:www.bury.gov.uk/culture.htm). As well as 'advertising' library services, the community information database is available, a 'bulletin board' style 'What's On' and 'What Do You Think?' feedback facility and enquiry button. Lack of staff time and problems getting access to change site information on the tightly controlled corporate site have hampered further development.

The management of change

The succession of initiatives that have followed the *New Library* report can seem bewildering, especially when added to other local authority, educational and cultural programmes: best value, modernisation, etc. Whilst these schemes clearly point in roughly the same direction, and meet the needs of New Labour and its perception of the needs of society at large, what about the needs of public libraries as organisations? If the implications of these changes are not managed properly, and the needs of public libraries themselves are not met, they will be in a weak position to effectively deliver the People's Network.

The management of change is a major theme of Management Studies. It tries to answer the question of why some organisations change with the times and go from strength to strength, whilst others either stick rigidly to the old ways and die, or attempt to change but it all goes horribly wrong. We probably all know of examples of the change that has not worked, the computer that sits gathering dust in the corner because no one knows what it is for or how to use it (the enthusiast who installed it left for a better job long ago).

There are several interesting and relevant areas of the management of change, including the stages that are gone through, analysis of forces for and against change and the role of 'leadership'. This article will focus on one area, the effect of introducing one element of change on all the other elements of that organisation. One of the most significant authors in this field is H.J. Leavitt [5], who identifies four areas of an organisation which are subject to change:



Task refers to the product or service that is produced, and the processes that are undertaken.

Technology is the equipment that is used.

People is fairly obvious, but includes the beliefs and ways of doing things (paradigms) that the staff follow.

Structure means the structure of the organisation.

Leavitt argues that each of these elements is interrelated, so that to affect one is to affect them all. Changing one element will have inevitable, and often unpredictable, effects on the other factors. Change the technology and you may end up changing the task (e.g. the effect of the invention of television on the BBC's 'task'), so you will need different People and a new Structure.

A brief look at the past introduction of new technology in libraries illustrates some of the points. Automated circulation and cataloguing systems have been around for long enough for their effects on other aspects of the library organisation to be observed. The much heralded freeing of staff from mundane filing has happened, though there are still routine tasks such as printing and sending out overdue letters. It changed tasks but perhaps not the end product of service to the public. Limited structural change has taken place as a new breed of systems manager, or some such, has arisen. On the other hand, cataloguing seems to have suffered in most public library authorities, though perhaps this would have happened anyway. It is harder to decide if the 'people factor' has changed. Library operations had to be more uniform throughout an authority, leading to less parochialism.

Other factors have intervened: more complex categories of materials (multiple AV types), more enquiries, new styles of management, coping with the failings of said computer system, etc. It is hard to separate the changes arising from automated circulation systems from other changes taking place at the same time.

Does this analysis help us understand what is happening in public libraries now, as new ICT is introduced? What effects might this have on the Task, People and Structure? If we can understand this, it may help manage the problems/opportunities the People's Network will bring.

Task and people

The services offered by the New Library have already been described: information, support for learning and reader development. The *New Library* report emphasises that these are things public libraries already do, but they will be doing them using ICT. However, this 'more of the same' is not to say that public libraries are not facing some revolutionary changes. The enhancements to services proposed are major ones in their cumulative effect, such that they do represent a potentially very significant shift of emphasis. In 2000 most public library staff, especially in the smaller branch libraries, spend most of their time issuing and returning books (the 'core service'). The New Library of 2010 should see most staff doing very different things in very different ways. They will be helping customers use computers in a support role for information, reading or learning.

Absent from most of the published literature on the New Library is the distinction between different types of public library. Yet there are great differences between a 'central' library in a large town or city, and the small suburban or village 'branch' library. This can be an official policy, represented by describing branch libraries as 'community libraries'.

Central libraries provide more in-depth and wide-ranging services, and their users include more business and higher education users. They are often divided into different service or subject sections, usually with their own staffing which effectively function on a daily basis as self-contained units. Most central libraries have distinctively different information or reference services and lending libraries, with separate staffing. Most professionally qualified staff are usually based in central libraries, or sometimes county headquarters. The culture of central libraries tends to emphasise quality of professional service, and they can be impersonal and anonymous (which many users prefer). Central libraries have tended to get most of the ICT resources so far. They have the specialist information services and professional staff that have most use for them. As they are the busiest service points and attract younger, better-educated higher education users, they will also generate most use. We can already see many elements of the 'New Library' in many central libraries all over the country, with Croydon being perhaps the best known example.

Branch libraries are often quite different, and much more varied according to local circumstances (particularly the priorities of the staff in that branch). The vast majority of their business involves the lending of books, with a strong bias towards popular fiction. Of course, lending figures have been in slow decline now for very many years and this is a cause for concern, especially in times of financial cuts (which have not gone away entirely) and an emphasis on quantifiable performance measures. Other activities include children's story times, coffee mornings, etc. Their users tend to be the local older residents together with mothers and children. The highest issue figures are for adult fiction and children's books. So, whilst information enquiry figures are generally low, the potential for reader development for adults and learning support for children is high.

Many of the staff will be part-time and most branch libraries have no professionally qualified staff based there. Staff culture is often highly social, and this social role is very valuable in its own right. This staff culture may be eminently compatible with the information enquiry, reader development and learning support work envisaged by the People's Network. In some ways, it might work better than the less personal approach of central libraries. For example, in the learning support role, where the user's whole life is affected and they will benefit from a less formal atmosphere. On the other hand, the existing cultures certainly feel different, and it is likely that, if they are performing different roles, they will be different.

It is here, amongst the large number of branch libraries, that the impact of the New Library will be felt most. True, take-up of ICT in smaller branch libraries, given their user profile described above, tends to be low, but it is clearly the intent of Government that they be geared-up for this new role. There is certainly a case to be made for providing ICT services in this context, physically close to where users live, especially the young, old and disadvantaged. They also generally provide a less threatening environment. Given the tremendous threats to branch libraries in recent years, and the on-going decline in 'traditional' book lending, perhaps this is just what they need to rejuvenate them.

Structure

Change in the internal structure of the organisation will take longer to happen. Will front-line staff with higher ICT skills also undertaking information/learning and reader support roles demand higher salaries and status? Will this apply to all front-line staff or just the increasingly important paraprofessionals? What will happen to management structures?

External structural change may take place, such as the blurring of some divisions between education and public libraries. Colleges are increasingly interested in providing community access facilities. Libraries have the capability to offer online learning from even the smallest branch. The voluntary sector is creating its own services. The opportunities for networking have never been greater, but public libraries are not the only ones in the community access to ICT game. What kind of structural change will emerge from this confused and potentially competitive market place remains to be seen.

Conclusions?

Much management of change writing emphasises the complexity of modern organisations. They may be so complex that they are, in some ways, unknowable and managers rely on 'instinct' to engage in 'mess management'. It is difficult enough to analyse organisations in their static state. So, it is even more difficult to predict the implications of changing one or more elements, doubly so as the introduction of ICT is one of several changes in public libraries and their environments. If this is the case, then there can be no conclusions.

Yet if anyone does know the environment of public libraries, and who can, therefore, make predictions to enable the implications of change to be positively managed rather than merely survived, it is public librarians themselves. Despite being subject to the direction and influence of central government, only public librarians can effectively manage change in their own organisations. History will tell us whether they were up to it.

References

1. Library & Information Commission. *New library: the people's network*. 1997.

2. Department for Culture, Media & Sport. *'New library: the people's network'. The Government's response*. Cm. 3887, 1998.

3. Library & Information Commission. *Building the new library: a report to government*. 1998.

4. Jones, Bronwen *et al*. *Staff in the new library: skill needs and learning choices*. British Library Research & Innovation Report 152, 1999.

5. Leavitt, H.J. Various publications including: *The social science of organisations: four perspectives*. Prentice-Hall, 1963. *Managerial behaviour*. 5th edition Chicago: University of Chicago Press, 1988.

Facilitating scientific communication for industrial innovation

Paul M. Evans
Elsevier Advanced Technology, The Boulevard, Langford Lane, Kidlington, Oxford
p.evans@elsevier.co.uk

Much has been written about how methods of working and communicating can improve the productivity of innovation for industry. Less has been related to this from the overall development of science and policies that assist this. The changing organisational context of industrial research brings the need for scientific publishers to reinvent themselves for this market segment. Scientific communication, including one of its key functions, awareness, is examined and it is concluded that functions and processes in scientific communication may be organised more efficiently to increase the productivity of industrial research. The new context of virtual communities, exploiting the opportunities for interactivity, provides the organisational basis for introducing new methods for inculcating new approaches to knowledge management, for innovation in industry to occur more effectively. An approach to better understanding knowledge synthesis and the potential role of the publisher, as communications facilitator, is discussed.

Introduction

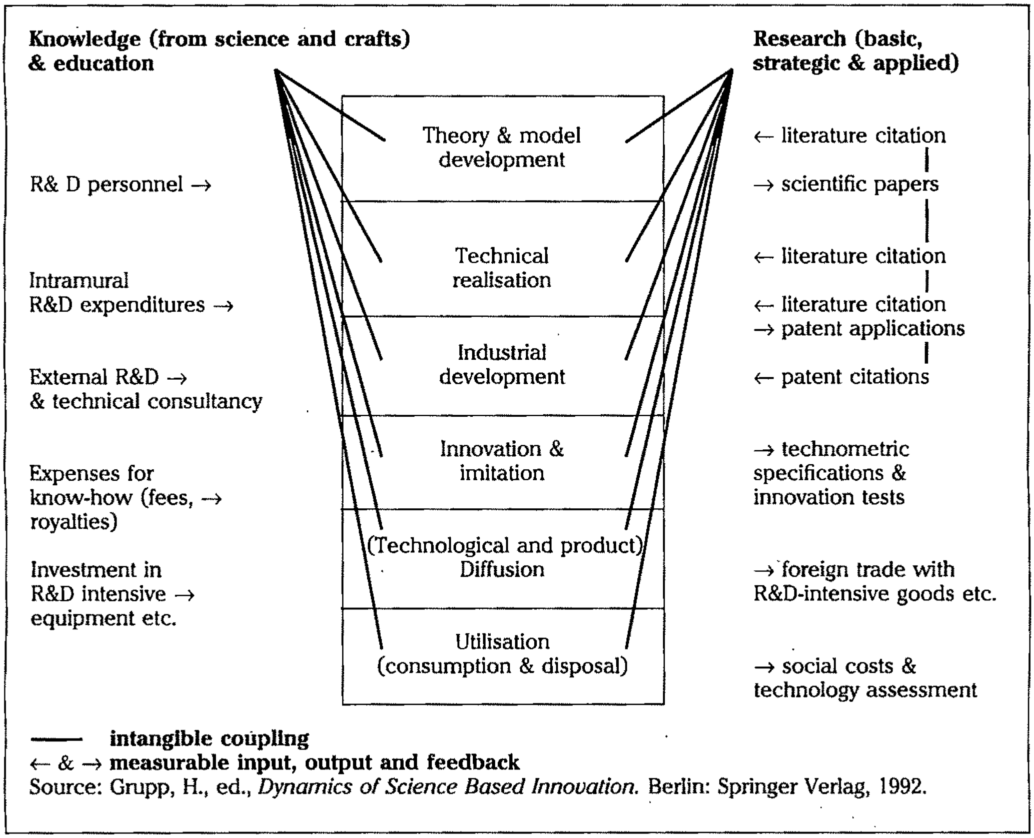
Much attention in government has centred on the influence of technological innovation to create, boost and form a new base for the modern industrial economy. A considerable amount of the study of innovation in management science has been centred on ways to recombine current technology, often in the light of better market information. This is innovation oriented around demand articulation and it became critical in an era when governments, such as the Thatcher Government in the UK, asserted that the market should itself be the main controller of the economy. Meanwhile, in social science more generally, innovation has often related to skills and practices, even ones that could be ritualised to ensure diffusion and widespread use. Hence organisation science has its roots in social science in the context of innovation studies. But it can be argued that the origin of all innovation is founded on the new knowledge arising from science, the art of discovery in the natural world.

This view separating human endeavour, a social activity, from the phenomena of the natural world (from the stars to sub-atomic levels) has been the base of what has become termed the natural science paradigm and its methodology, a hypothetico-deductive model, the scientific method.

When applied to innovation, itself the product of social interaction and science as a base, the models arising suggest a mainly one way flow from science to innovation and on to diffusion and utilisation. One more detailed example of such a model is made by Grupp, Albrecht and Koschatzky [1] (see Figure 1).

The model identifies one sphere of influence on innovation being related to the activities of knowledge from research and development, (representing interplay between origins in both science and crafts and situated mainly in industry). The other sphere of influence is research itself, perhaps a more erudite activity related more to the academic arena where scientific papers are the key building blocks and markers of progress. As these two spheres interact there can be fruitful collaboration in the development of theory and models, in developing technical outputs and then ones appropriate for industry before leading to specific product development, diffusion and utilisation. However, this could be seen as representing a chain of activities akin to theories of tech-

Figure 1. *A schematic model of the coupled functions of science, technology and innovation.*


Knowledge (from science and crafts) & education
Research (basic, strategic & applied)
R& D personnel →
Intramural R&D expenditures →
External R&D → & technical consultancy
Expenses for know-how (fees, → royalties)
Investment in R&D intensive → equipment etc.
Theory & model development
Technical realisation
Industrial development
Innovation & imitation
(Technological and product) Diffusion
Utilisation (consumption & disposal)
← literature citation
→ scientific papers
← literature citation
← literature citation
→ patent applications
← patent citations
→ technometric specifications & innovation tests
→ foreign trade with R&D-intensive goods etc.
→ social costs & technology assessment
—— **intangible coupling**
← & → **measurable input, output and feedback**

Source: Grupp, H., ed., *Dynamics of Science Based Innovation.* Berlin: Springer Verlag, 1992.

nology push from science to industry as was common in early models of the innovation process. Although Grupp does not specifically limit these processes to representing a linear chain that begins with science there is a risk for many less experienced in innovation theory viewing the model in such a way.

What is less clear in an otherwise well defined model is the influence of interaction (e.g. between organisational influences and technological ones), especially between layers, which may not be adjoining. It gives the impression of a more particulate, specialised process than is necessarily the case, especially in more applied areas. Central organising principles may be held in a sense of vision in the applied area throughout the process, emphasising the output orientation. This shared vision across the process may be best supported by good communication across and between the various levels. It is also likely to lead to increased productivity through better targeting of the research effort towards desired outputs. However, the assumption that this will be so needs to be tested in practice via experimental participation with researchers in the applied science and technology areas. But it does suggest that rather than representing a chain of processes the movement from research to innovation is more akin to a pathway and an analogy can be made with the complex biochemical pathways for respiration in the cell, releasing energy.

What is clear in the model by Grupp *et al.* [1] is that two provinces of original value are represented: (a) knowledge (embodied as a social phenomenon) and (b) research (conceived as a scientific phenomenon). Analysis of related processes requires these twin dimensions to be carefully considered. Another pair of interest is the industry/science couple. The culture and values of each are distinct. On the one hand, science is founded on shared knowledge and ultimately public information. On the other, industry relies greatly on its ownership or competitive advantage in intellectual property and reflects commercial values. At times the distinction in values may inhibit progress from research through to innovation and certainly communication between specialists involved depends on clearly understood policy and practice with regard to intellectual property rights as discussed by Leer [2, ch. 3].

Even though universities are becoming more commercial and more concerned with intellectual property created by staff and also with related rights, there is likely to remain a distinction in culture and values that may seem to set this world of research apart from the related world of innovation. Meanwhile the mechanism of more devolved research industry does emphasise communication aspects and the significance of networks in creating new value and exploiting new knowledge via applied and strategic research.

In the scientific publishing industry these trends have particular significance especially in the era of apparently maturing product lines of paper journals, and the rise of new opportunities at the desktop of individual researchers via the Internet. In aiming to establish new forms of business, the task of improving communications in order to increase productivity in research could be particularly relevant to industrial areas. It may even suggest the opportunity of an extended market towards high-end technologists in firms concerned with new product development based on high technology. Referring once more to the model by Grupp *et al.* [1], the meeting place of links from knowledge and research is not simply a transfer of information but an engineering process resulting in the synthesis of new knowledge. For this to be successful, scientific publishers have to be close to content but not necessarily controlling its manufacture. The task is rather to empower the participants in the knowledge synthesis processes to inter-relate aims and new findings via appropriately maintained communication channels, information archiving, validating and related quality management processes using technology to connect related people, concepts and problems. In essence the publisher becomes a communications facilitator, according to Roosendaal [3], assisting the synthesis of new knowledge for the applicability of science and technology.

Knowledge synthesis can be viewed to an extent as a branch of organisation science. This is because, while knowledge management is becoming a central organising principle for the new organisation in the knowledge-based economy, the synthesis of that knowledge and its effective utilisation is critical for adding new value and achieving competitive advantage.

Knowledge synthesis underpins the new economy for a new century as there has been a realisation that management has focused its attention on a minority portion of value in firms in terms of conventional business assets, rather than those often written down as representing good will when acquisitions occur. Prices paid in acquisitions indicate a higher value is placed on systems, skills and core competencies whereby the new knowledge for firms to adapt and exploit their changing environment can be synthesised. This realisation has been one of the factors behind the rise of new organisational models such as the KBNO, or Knowledge Based Network Organisation. The model has come into being because of:

- the general increase in education in the latter part of the twentieth century;
- the global and unprecedented move towards democracy;
- the search for a merger of the proven qualities of capitalism (innovation particularly) and more collectivist expectations;
- the emergence of especially free subcultures particularly related to the developments of the Web communities using the Internet.

These aspects are described by Frederic Voulouzan, one of the international MBA students at Nijenrode University, the Netherlands

School of Business, in a research pamphlet [4], an output of a collaborative group led by Annemieke Roobeek and Erik Mandersloot.

The following year's session's output at Nijenrode on KBNOs is contained in another pamphlet, *KBNO – Networks for the third millennium* [5], where organisational aspects such as the mobilisation of collective intelligence and team-based empowerment are presented in the setting of dominant factors. These dominant factors include the transfer of power from governments and corporate businesses to stakeholders at all levels of society, as well as the growing understanding of the need for flexibility and learning to deal with the unpredictable business environment. Most significantly KBNOs move across and beyond the boundaries of the firm and lead to a high level of virtuality. One example is the 'Vision Web', which is based on eight different organisations in the information communications technology arena. Rather than just acting as a web page or chatbox, Roobeek describes how the KBNO form mirrors the Dutch 'Grand Café' concept, a café/social setting where people meet each other and exchange ideas. Customers and other consultant participants join together to form a social setting where there is trust and bonding between network participants and, most importantly, a focus on the creation of new ideas and innovations.

For the publisher seeking to be active in the field of industrial high-tech innovation this model of the KBNO becomes increasingly important as networks of knowledge professionals at universities, in science parks and within leading innovative corporates intertwine further and traditional boundaries become blurred. Interactivity using the Internet technology is a vital component of this transformation and publishers have to address how they can transform themselves rather than merely automating their previous paper-centric basis as information providers.

Technological/scientific progress and economic change: the significance of new knowledge

For the scientific publisher the key interest here is how changes have been translated into the market for their activities as publishers. The publishers are beginning to reposition to not simply record the formal processes of official publication, but to feed a new breed of knowledge workers via the development and maintenance of informal communications channels. What is the basis of this change?

First, there is the economic context to technology which has a new significance in the way activities are organised and prioritised. Studies on innovation made in the 1980s, by Rothwell and Wissema [6] among others, referred back at times to theories of there being long waves in economic activity based around technology eras. The best known of these theories was developed by N. Kondratieff, a Russian economist from the time of Stalin, who believed he could define long cycles of economic activity lasting around 52 years. Each cycle effectively began in a recession of the previous technology being displaced and there was then a ferment of ideas and development leading eventually to a period of growth and regeneration lasting for about 25 years. The first Kondratieff cycle was thought to be caused by the introduction of textile machinery and the application of steam power. Later cycles were associated with railways and mechanical engineering (2nd long wave); chemistry, electricity and internal combustion engines (3rd long wave); and, subsequently (after Kondratieff) the fourth wave has been associated with electronics, aerospace and chemistry, and the fifth (current wave) from the 1970s onwards with microelectronics/computing and biotechnology.

If we apply Kondratieff's theories we are now just a handful of years into a period of regeneration of about a quarter of a century, in which there will be the economic benefit of a culmination of the effects of a range of new products related to the new dominant technology. The new dominant technology probably arises from the enabling functions of computer technology as we move forward in what was originally termed an information age, but which could now be termed a communications age and may conceivably be termed a knowledge age [2, 7]. However, Grupp, Albrecht and Koschatzky [1] assert that the precise starting point of future cycle periods using Kondratieff's waves is not as useful as H. J. Gerster [8] used filter theory to show that waves actually varied with periods of between 20 and 60 years or so. In fact, apart from Gerster's work, there are a number of controversies related to Kondratieff's theories. Firstly, as Gerster indicates, statistically it is difficult to prove that the waves exist even in regard to measuring world economic activity over the years; secondly the causation of the waves is a difficult theoretical area. The most prevalent views follow on from the work of the Austrian economist Schumpeter who believed innovations resulted in changes in the underlying 'techno-economic paradigm', as this was to become known. Thirdly, there is dispute over whether major innovations are made during the period of decline of the last wave, or during the upsurge in prosperity of the next one, so causal links are again complex. Indeed, some argue it is the anticipation of the economic upturn that leads the reversal from decline rather than the innovations themselves.

Whatever the case it is clear that in these cycles of the economy and technology/innovation, there is also a strong interplay with the cycles related to social and institutional structures, such as government and all its forms, as well as the establishment of new industrial firms.

Even if the periods are not uniform it is not necessary to predict starting points of cycles. Surely, what matters is to know that there are these periods and whether you are in the ascending or descending phases. The general view that technological innovation enters intensive periods that have a major effect on the economy is well founded, as is the view that we are now in such a period of development with a special significance for computer communications technology.

Of course, technology and its development do not in reality exist outside the social context. The contrary view, known as technological determinism, is to some extent refuted by the fact that governments now actively pursue industrial strategy and establish technology policies to ensure development does actually occur. In the 1980s the contrast was marked between, for example, on the one hand the UK with its market-dominated policies (one form of strategy) and other strategies in other older industrial nations supporting sunset industries, and on the other the high growth economies of Asia (at that time) and part of the USA with government support focused on sunrise industries. This new type of strategy has put an emphasis on what are known as generic technologies (or critical enabling technologies) and therefore upon applied research and strategic research. The necessity is to follow the waves of technology and try to discern what the sunrise industries are going to be.

Competitive advantage at the level of the nation state (perhaps itself already something of an anachronism in the global economy) depends on the development of intellectual capital via investment in training, education and research. Drucker [7] assesses this at around 20 per cent of GNP for the US economy and notes that this a higher proportion than occurs in the form of money capital investments in the economic future. In fact he predicts that in economics there will come a major intellect to provide a new theory based on knowledge as the source of value, perhaps writing *The Knowledge* to finally outdate Marx's *Das Kapital.* Others at this time are setting technology as a prime source of value rather than a derived one and that it has the same validity as labour, capital and land have had in traditional economics.

Even beyond the seemingly rational world of economic theory, Kodama [9] asserts that research has a new significance, especially in Japanese firms, where investment in research has overtaken significant ratios with investment in manufacturing plant and even economics based on pay-back principles. Instead research, in his view, has become an environment in which it is necessary to keep up just to be able to compete. It is like a membership fee for being able to operate as a high-tech company.

Kodama presents the changes as a new techno-socio paradigm. As an alternative one can see this as new ways of fulfilling familiar functions, especially in the case of publishers. Roosendaal's proposition [10] is that for publishers the transformation of products and services will result in the fulfilment of these functions and the needs of researchers (although in new ways). Any effort to improve the productivity of research in terms of short term or long term outputs should have beneficial economic effects. Communications within science and related innovation processes seem an area of timely concern on which to focus to take advantage of new computer network technology.

This is why scientific publishers are looking afresh at the market and the knowledge workers they seek to serve on the overlapping boundary between the worlds of science and industry. They are thus focusing more on the reader, the user, and the professional rather than purely on the library. This requires thinking again about the institutional set-up as publisher. Instead of focusing on the existing client, the library, they need to look at the client of the client and how they can assist by working to speed up the innovation process. This requires a closer relationship with the end-user market and a better understanding of what is needed and in what form. This logic underlies in part the reasons for the acquisitions in 1998 by the science publisher Elsevier Science of web-based community publishers, such as Engineering Information (with its Engineering Village), Biomednet and Chemweb.

Scientific communication, its functions and knowledge synthesis

From a traditional viewpoint STM (scientific, technical and medical) publishers reproduce papers by scientists usually within journals after a period of peer review and amendment. Taking into account the production operations this may well take anything from six months to well over a year. The work is thus often produced after the main scientific or knowledge synthesis process has occurred. In fast moving fields it is often the case that related researchers in the same field read preprints of the paper (a pre-publication version without amendments requested by referees) and, often nowadays, this is posted on the Internet at the researcher's institution. Thus the publisher is mainly engaged in validating or registering the work so that it can be distributed for archiving.

Since more and more science is being published, but generally scientific library budgets are under considerable pressure, it is hard for publishers to maintain their subscription base from year to year. This has led to the perception that the traditional paper journal represents a maturing product line. As a result STM publishers are investing heavily in computer technology to develop new lines based on the Internet. In the transition phase, operational costs of journals are quite large as both paper publishing and electronic publishing functions are supported, often independently. Leer [2] criticises Nicholas Negraponte, the Director of the MIT Media Lab, for his view that the paper journal will disappear to be replaced by electronic journals. Certainly it seems clear that in new periods of technology older forms persist for a long time. What is more, the opportunity for transformation in supporting communication for a community of users and exploitation of the interactivity of the new medium is a positive development. To miss this is to risk simply automating the paper journal into an electronic form that represents a more negative development.

This new context for scientific communication could be thought to represent a new paradigm, and related in a way to the apparent new paradigm for research described by Kodama in the age of technology fusion rather than technology breakthroughs. Roosendaal [10] however proposes that, to analyse the discontinuities, which may occur in the forms of products and services, it is best to focus on the key functions of scientific communication already established and see how they may operate in a new context. Hence he defines four main functions: registration and archiving, described above, which are mainly external to the scientist and hence objective; and the other two are certification (the review process to determine and maintain quality) and awareness, which are internal to the scientific community and the researcher. Aspects of these last two could be organised in a more external fashion via the publisher using technology to add value to the process. The aim of adding value to the communication process would be to increase the productivity of research processes and applied outputs.

The market for scientific communication is described by Roosendaal in terms of four main forces and their interplay:

- the actor pair (author/reader) – where the need of registration and certification of new findings meet the need for fulfilling awareness in the wider community and applying the benefits of scientific output;
- accessibility – where electronic delivery can play a significant role;
- content – where quality management and judgements concerning relevance to end-users begin to fuse in the applied context of industry;
- and applicability – which is the key criterion of concern in this paper.

In effect, the context surrounding research in terms of policy and the acceleration of outputs is highlighted in the shift of the publisher's role, from that with a product or production process-oriented emphasis based on the provision of information, to that of a marketing or relational one based on facilitating communication. The function of awareness gains in this transition. This is significant, as the phenomenon of technology scanning is critical to industrial progress.

The stress on awareness and applicability emphasises that for genuine new innovation the knowledge synthesis and exchange is, above all, an active process. The virtue of the new web environment for the organisation is that it is an interactive environment and this aspect will become increasingly important.

It is in fact quite possible that the new direction for interactivity in the synthesis of knowledge will be spearheaded by universities. In recent years educational technology has turned from a dependency on mass communication methods, such as lectures, to increased use of group work and participation in order to inculcate higher cognitive skills than were often achieved with the traditional but cost effective methods of the past. Technology enables some of the costs to come down through, for example, delivery to wider audiences, such as via the Internet, as the burgeoning number of distance learning courses now on offer testifies.

It is also interesting to see how universities are changing their role and facilitating the growth of intellectual networks tied more closely to innovation processes. In Sweden there has been a move to set up alongside the traditional aims of education and research within universities, a third aim of development, particularly in regional settings at new universities. The emphasis is now on the interaction with practitioners, as discussed by Brulin [11], replacing the previous model founded on noble seclusion and the 'dream of High Science' in the German tradition and following Plato and Cartesian philosophy with the emphasis on theory rather than practice. But valid knowledge has now a closer relationship with practice. This is especially so in social science research with the link to development efforts in working life involving practitioners outside universities, gives support to the foundation of knowledge based network organisations. At the same time it accompanies a democratisation of the research process as with the so-called Learning Network Program, which also involves Swedish labour unions.

This emphasis on participation is mirrored in the establishment of such organisations as the Network University, which is associated with the University of Amsterdam. Here the Internet is enabling participants to work in groups remotely and to build on the knowledge within their network. Students can:

1. communicate with each other;
2. gain direct access to a post-doctoral student via an academic call centre;
3. gain access to the university professor who developed the academic programme;
4. interact with the programme and feedback into it to enrich, expand and improve the programme and the knowledge.

The supervisor can also follow the student's trajectory through the course by examining an electronic workbook that is kept as a record. Courses so far include one on the financial crisis in Asia and one on the introduction of the Euro.

For publishers a model is being drawn to orchestrate interactivity in knowledge synthesis much as the programmes on such courses do. What is more, the Swedish experience, developing interaction with practitioners, provides an opportunity to extend the end-user or participant base into the realm of knowledge based network organisations involving universities, high-tech company practitioners and consultants. It also leads the way from supporting research alone to the combination of research and development, hence innovation in business.

Meanwhile, as the publishers seek to find ways to transform rather than automate their function in tackling the issue of interactivity, they are addressing their customer needs in seeking wide awareness of information concerning the environment they work in, such as new technological developments, economic developments and legal changes.

Everett Rogers is a key academic figure in the field of technology diffusion and his observation of awareness as a function is that it is not actually passive but active [12]. This is all the more apparent nowadays in the context of the World Wide Web where intelligent agents are being developed to improve regular, intermittent searching and scanning of external information by knowledge workers [13].

But Drucker [7] in categorising knowledge differentiates between fundamentally new knowledge, which is closer to science, and other forms which relate to improvement and exploitation (i.e. diffusion) and really concern making further use of the available knowledge base. For him fundamentally new knowledge is 'genuine innovation'. However, the word innovation is used differently by different writers and researchers so that it often encompasses exploitation of existing knowledge. This places a special significance on new links to basic research as well as applied areas and presupposes that the global society must make a special effort to continue investing in the intellectual capital of basic science, even if the outputs are long term and hard to measure.

The changing world of scientific communication has to take account of the increased emphasis on applied research and strategic research in arriving at models for the electronic age, increasingly involving industry's use of contract research organisations (CRO's). However, it is encouraging to note that, even disregarding the potential for links outside an inner scientific circle, there can be productivity gains arising from computer mediated communications for scientists by allowing researchers ready access to like-minded colleagues who happen to be at another institution. It also enables young researchers at less prestigious institutions to engage in dialogue with more experienced researchers at centres of excellence as Walsh notes [14].

In the industry arena Leonard-Barton [15] has a model of knowledge creation and diffusion which pinpoints four areas of activity in the firm where knowledge is synthesised or transferred: knowledge importing, shared problem solving, implementation and experimentation. This cyclic model can be usefully applied to the area of industrial research but overlaid on the network connecting related scientists, industrial firms, universities, publishers and so on, rather than just the individual firm.

The publishers seek to import knowledge into the firm to assist innovation and that is the current state of play. In the future the aim may be to orchestrate the interactions within the firm more effectively via, for example, incorporation within intranets and even their extranets. This is a fundamental shift of position for the scientific publishers in the industrial environment.

The context of knowledge management in the social sciences, including management literature on innovation, overlaps that concerned with research management in the sciences and the pursuit of increased productivity.

Technology management

While sharpening the focus of innovative communication for research toward scientific content it is important to note the lessons of what has been called 'third generation research' (as for example by Rousell, Sade and Erickson [16]) with R&D organisations better integrated 'spiritually and strategically as full and equal partners in the business enterprises whose prosperity they were intended to serve'. The flow from research to development and then production is no longer sequential in practice or conceptually.

The success of Japan, especially through the 1970s and 1980s, is based less on the input of scientific theory and models than on the ability to manage the applicability and processes of application. As Sony's CEO, Akio Morita, puts it 'Technological management will be the key success for companies anywhere in the world in the coming years'. This perspective is important in assessing the boundaries and focus for communications in applied research. It is emphasised by Nonaka and Takeuchi [17] and seen to be particularly successful in the Japanese context of middle-up-down management. The organisational movement of knowledge depends first on its synthesis and transformation. There is often a personal transfer of information and this is organised within interdisciplinary teams. For the publisher the significance is that content must be available to people from different professional and technological disciplines in order to have full applicability.

Different forms of interaction between tacit and explicit knowledge can result in knowledge synthesis. The most critical steps involve 'articulation' (from tacit to explicit) and 'internalisation' (from explicit to tacit). While these observations are based on corporate knowledge management, they can be applied more generally to the area of applied research, and include such technologies as computer supported collaborative working (including groupware tools), distributed hypertext systems, document management systems, intranets and various tools assisting knowledge representation, e.g. semantic networks and concept mapping of knowledge objects. These are described by Mentzas [18] as essential for the new age of the knowledge-leveraged enterprise. They have particular relevance for the scientific publisher serving researchers close to the field of innovation and seeking new opportunities to increase productivity.

It can be argued that Nonaka and Takeuchi tend to under-emphasise organisational institutional barriers and are quick to criticise those who seek to tackle these areas, such as Peter Senge who employs a systems approach in developing the theme of the 'learning organisation'. In fact, a more integrative approach is required to deal satisfactorily with methods to increase communication for the synthesis of knowledge. Engineering perspectives tend at times to underestimate the impact of barriers which include strong organisational biases that are not usually formally acknowledged [19]. Publishers have to take account of this new terrain in plotting a course for the future to emphasise applicability, relevance and awareness while maintaining quality through support of certification and the development of a unified scientific archive.

However, the limitations of corporate research globally have been partly due to the fact that, in its third generation form, most of the attention has been on links within the firm and not enough on links between organisations, including universities and even competitors in the same industry. This background to the fuller organisational environment could input into the communication channels for research and innovation in a more co-operative framework of mutual benefit to the organisations involved. But it would depend at critical stages in competition on the ability to clearly establish and protect intellectual property rights within the communications process. The boundaries, where co-operation and competition could conflict, require research and the maintenance of appropriate trust within the network of participants close to the research.

This is where the publisher can pursue a role at the hub of the 'knowledge based network organization (KBNO)'. The KBNO is a useful model to consider for the new industrial and organisational context of the market in which the publisher is beginning to operate. This context is now the critical one for the scientific publisher to assess in repositioning to take care of new needs in the industrial markets served.

Conclusions

The transformation of the scientific publishing industry in taking more clearly the role of scientific communications facilitator can only be successfully monitored and achieved by paying attention to key needs and functions satisfied in the past and in the future. This framework of continuity for the analysis enables new products and services to be evaluated provided that the scientific research process is more clearly understood by publishers. In the case of research close to industry, where a key concern is on increasing productivity, relevant empirical data based on bibliometric analysis, qualitative data from interviews and resulting from partnerships with the market and collaborators, can combine to derive relevant theory and allow it to be applied in the context of a virtual organisation or team of related areas of research and industry. Most importantly publishers have to be prepared to reposition themselves in order to address end-user needs more directly and also more interactively in order to address not just awareness issues for their market but applicability and the synthesis of new applied knowledge close to science.

In undertaking such a field of enquiry and practice, publishers, and the market of which they form part, are able to integrate aspects of learning and research from research management, innovation and strategy theory, the sociology of science, organisational learning as well as technology associated with electronic publishing. The research is interdisciplinary and itself focused on its output as practical implementation. This represents a new sub-area and set of methodologies to be used in the merging field of knowledge management related specifically to knowledge synthesis, an area of prime economic importance. Meanwhile the active modelling of new tools for the end-user continues and experience of what works in practice will guide the future direction of science publishers serving industry.

References

1. Grupp, H., Albrecht, E. and Koschatzky, K. By way of introduction: alliances between science research and innovation research. In: Grupp, H., ed. *Dynamics of science based innovation.* Berlin: Springer Verlag, 1992.

2. Leer, A.C. *It's a wired world.* Oslo: Scandinavian University Press, 1996.

3. Roosendaal, H.E. Scientific communication and the Science Process. In: Van Luyt-Prinsen, J.G., Meijer, E. and Ticer, B.V., eds. *Proceedings of the International Summer School on the Digital Library.* The Netherlands, Tilburg: August 1996, p. 1–12.

4. Roobeek, A.J.M. and Mandersloot, E. with the international MBA class. *Connectivations: networking for innovations: International MBA Programme.* Breukelen: Nijenrode University, 1998.

5. Roobeek, A.J.M. and Mandersloot, E. with the International MBA class. *KBNO – Networks for the third millennium: seminar on knowledge based network organisations: International MBA Programme.* Breukelen: Nijenrode University, 1999.

6. Rothwell, R. and Wissema, H. Technology, culture and public policy. In: Rosenberg, G., ed., *Management of technological change.* Oxford: Elsevier Advanced Technology, 1991.

7. Drucker, P. *Postcapitalist society* New York: Harper Collins, 1993.

8. Gester, H.J. *Lange wellen wirtschaftlicher Entwicklung.* Frankfurt: Verlag Peter Lang, 1998.

9. Kodama, F. *Emerging patterns of innovation.* Boston: Harvard Business School Press, 1991.

10. Roosendaal, H.E. and Geurts, P.A.T.M. Forces and functions in scientific communication: an analysis of their interplay. In: *CRISP 97 (Co-operative research information systems in physics).* January 1998. Available at: http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html

11. Brulin, G. The new task of Swedish universities: knowledge formation in interactive co-operation with practitioners. *Concepts and Transformation,* 3(1–2). Amsterdam/Philadelphia: John Benjamins Publishing Company, 1998, 113–128.

12. Rogers, E.M. *Diffusion of innovations.* 4th edition. New York: Free Press, 1995.

13. Krulwich, B. *Intelligent agents for Internet automation.* At European ESPRIT IT Conference, Brussels, November 24–26, 1997.

14. Walsh, J.P. *Telescience: the effects of computer networks on scientific work*. At ESPRIT European IT Conference, Brussels, November 24–26, 1997.

15. Leonard-Barton, D. *Wellsprings of knowledge*. Boston: Harvard Business School Press, 1995.

16. Rousell, P.A., Saad, K.N. and Erickson, T.J. *Third generation R&D*. Boston: Harvard Business School Press, 1991.

17. Nonaka, I. and Takeuchi, H. *The knowledge creating company*. Oxford: Oxford University Press, 1995.

18. Mentzas, G. *Information technology infrastructure for enabling corporate knowledge management*. At ESPRIT European IT Conference, Brussels, November 24–26, 1997.

19. Argyris, C. *Knowledge for action*. San Francisco: Josse-Bass Publishers, 1993.

CONTENTS

163–165 Comment
The information 'player': a new concept for the information user
David Nicholas, Tom Dobrowolski, Paul Huntington and Jonathan Raper

166–173 Information sharing in an online community of journalists
David R. Millen and Susan M. Dray

174–184 'Hyped' text markup language. XML and the future of web markup
Claire Warwick and Elliott Pritchard

185–193 Embarking on the Internet: what motivates people?
Reijo Savolainen

Comment

The information 'player': a new concept for the information user[1]

David Nicholas, Tom Dobrowolski, Paul Huntington and Jonathan Raper
Internet Studies Research Group, Department of Information Science, City University, Northampton Square, London EC1V 0HB
Nicky@soi.city.ac.uk

We need a new vocabulary to help us understand what is going on in the new digital information environment in which we find ourselves. The vocabulary we currently employ represents another age, and is proving an impediment to understanding and functioning in this one. We are imprisoned in a Renaissance vocabulary yet work in a virtual world. New metaphors and a new language are needed. The problems this causes are only too obvious to those whose daily lives put them in contact with the Internet. Most of all we need a new vocabulary for evaluating networked services like the Web, Intranets and WAP mobile phones. Without it information systems will develop a life of their own, irrespective of the actual needs of the information consumer or player. Some would argue that this has already happened. The previous generations of information systems took little cognisance of the user, but that has to change as we move rapidly towards personalised information delivery, bespoke information services and the like. We have rapidly moved to a dynamic, complicated and sophisticated information environment and our terminology, practices and thinking need to reflect this.

You might ask what is in a name? The answer must be everything. Terminology drives our thinking and, most importantly, the perceptions of the people we serve. Consider, for instance, the extent of professional soul-searching associated with the term librarian[2].

Essentially the argument is that we need to change our words to change our thinking, and we need to start with what is probably the most central (key) word of them all – user [2]. The word user (and users), like information, has lost much of its meaning. It is a tired, over-used, cheap and misused word, which provides the information profession with a debased currency. It is used to describe both someone who uses an information service and someone who *may* use an information service. This can hardly be regarded as being precise or helpful. Also, it no longer reflects the close and complex engagement that takes place between a person and today's interactive information systems. The word users paints a picture of a featureless mass, a lumpen proletariat, a homogenous body – people who are accustomed to being fed information in batch-processing mode. It is too passive and too mechanical for today's information environment. It does not recognise the fact that digital information systems create much greater information diversity than print-based systems. It also fails to convey that today's 'user' has to be fleet of foot to succeed in these fast-changing times. Basically, it is the wrong word, in the wrong place, and at the wrong time.

---

[1] Based on the article *The Information Player* [1].

[2] See for instance Circle of State Librarians, Annual Study Conference, *Who am I?* Held on 7 February 2000, at Institution of Civil Engineers, London.

Too much time has been spent tidying up and tampering with the word user. The time has come to abandon it, or prescribe closely its meaning. What we really need is a more active and accurate word. *Information Player* could be that term. Player used in the context of sport – meaning a football or cricket player – or in the context of business – someone who 'plays' the financial markets. It is a term much richer in meaning, one that conveys action and individuality. It is a term that recognises that today's information seeking can be interactive, recreational, social and competitive.

- *Interactive.* Today's user plays a much more important, complicated, creative and engaged role in the information seeking process. The term user is one-dimensional and, if dynamic at all, this is only in a linear sense. Player on the other hand suggests a multidimensional evolving relationship – the kind of relationship that is so much a feature of modern day information systems, like the Web. A player is constantly looking for new routes to goal – evaluating options (information) as they go. The player is part of the system while user gives the impression that they stand outside the system. Players operate in an information space (cyberspace) while users use information systems. There is a good analogy to explain why a change of name is needed. Using traditional information systems was rather like going into an office canteen, at a specific time, asking for items on a restricted menu, having that given to you, you eating it and someone coming and taking away your plates. Today's information systems are more analogous to the buffet lunch you order from an outside food caterer. You choose what you like, when you like, you take it to the table and you choose whether to return the empty plate to the trolley or leave it on the table. This plainly represents a fundamental change in behaviour. In the latter case you are much more involved and in control of the whole process. Determining the success or otherwise of the function – in this case the provision of food, but in our case the provision of information – will require us to examine different things. That is the point that is being made.
- *Recreational.* Today's information systems are now very much an extension of our real life – and this will prove to be even more so as the digital mobile phone expands its horizons into the information retrieval domain. The word player comes from the real world, whereas the term user is more of a theoretical construction. Information seeking today is no longer just about professional problem solving: it can be recreational (and mindless) too[3]. Indeed, the distinction between the two in the serendipitous and supermarket environment of the Internet is increasingly blurred. In a sense this has been partly recognised through the use of the term surfing (for information), but this word is rather too shallow a description for the process.
- *Social.* Information gathering is a social activity. The social context of the term is very important. Players play with somebody else. Players have different social roles: you have, for instance, teacher and student, politician and voter, journalist and reader.
- *Competitive.* There are costs – financial and/or time costs, associated with finding – or not finding – information. You can win or lose in the information locating game. Players can be seen as information investors.

The word player has other important connotations. Firstly, it is a term that acknowledges the new economic and political realities. There has been a shift in power from information producer to information consumer. The information consumer now holds centre stage. This very fact worries academic librarians and Media moguls alike. Indeed, anyone who manages large, centralised, inflexible batch-processing style information factories. Today's consumers have a wide choice and can quickly vote with their feet (mice). As a patron of a university library you were always supplicants – the Inter-library

---

[3] Research has shown that peak use of web sites is at office lunch times [3].

Loans (ILL) librarian had particular powers to open or not the information gates. Now the academic can go to the Web – Amazon.com etc. The ILL librarian will have to undercut the opposition now: the tables have turned.

Secondly, the word player is also closely connected with the term spectacle. This is important because the Internet is the biggest capitalistic spectacle of our times. Witness how people pay fortunes for impressive Internet addresses. Witness the amazing rise of e-commerce. There are fortunes to be had for the big players and games and adventures for the not so lucky. We are both actors in, and spectators of, the spectacle.

Thirdly, the term player is very much an Internet-type word. The Internet is very much part of the liberal world economy – the word user is most certainly not, but player most definitely is. There is a pressing need to get our words in line with the vocabulary of the Internet. The Internet has its own rich and picturesque language for describing itself. We cannot ignore this. It is only by employing this language that information professionals can address the huge information audience that the Internet commands. We are dealing with post-Modern reality. The first goal of a new language is to facilitate narrative discourse. In a post-Modern world it is discourse which creates and legalises vocabulary. Discourse also constitutes a research activity.

The concept of a player can be helpful in understanding how people interact with information systems – and re-visiting what we have said and learnt about them in the past. Continuing the sports analogy, few goals are probably scored in a textbook manner. In any football game players do a lot of things that are not in the training/coaching manuals. Similarly, a lot of information is collected by unconventional or serendipitous means. Thus, maybe, what we saw as minimalist and idiosyncratic information behaviour was not so odd, strange, etc., after all. Using the player concept helps us to get a better understanding of what the logs – and what we, witness at the terminal, PC, mobile phone, etc.

Finally, perhaps the profession’s notorious neglect of the user can be explained in some way by the term itself. For after all such a general and vague term hardly reminds us of the primacy of the individual and the necessity to investigate the individual’s information needs. The term player demands such an investigation. It sends all the right signals. And that is what the Web is all about – sending the right signals.

References

1. Nicholas, D., Dobrowolski, T. and Huntingdon, P. The information player. In: Scammell, A., ed. *Handbook of special librarianship and information work.* 8th edition. London: Aslib, (to be published Autumn 2000).

2. Julien, H. Constructing ‘users’ in library and information science. *Aslib Proceedings*, 51(6), June 1999, 206–209.

3. Nicholas, D. and Huntington, P. Who uses web newspapers, how much and for what? A log analysis of The Times/Sunday Times web sites. *NetMedia99 Conference: Proceedings.* London: City University, 1999.

Information sharing in an online community of journalists

David R. Millen
AT&T Labs, 100 Schulz Drive Red Bank, New Jersey 07701 USA +1 732 345-3396
drmillen@att.com

Susan M. Dray
Dray & Associates, 2007 Kenwood Parkway, Minneapolis, Minnesota 55405 USA+1 612 377-1980
dray@acm.org

Informal groups of professional workers are forming online communities to create and share information with each other. This paper offers an analysis of an online community of news reporters. Through participant observation in the group electronic discussion, and analysis of archival copies of the group conversation, the authors explore the ways in which this group requests and offers assistance to each other. The types of information created are examined and the temporal characteristics of the group are explored. The analysis revealed a vibrant online community of professionals who use the online communication medium to compliment professional face to face meetings and other forms of communication such as newsletters and journals.

Introduction

In recent years, there has been considerable research interest in online or virtual communities. One question of considerable interest has been the nature of the online interaction or the 'degree of community' that is developed and maintained by the online group members. It has been suggested, for example, by Dray and Siegel [1] that online gatherings fall along a continuum. At one end are transient chat rooms and temporary discussion groups in which little or no sense of community is ever established. At the other end of the continuum, are groups that grow to exhibit many community traits including 'multidimensionality, constraints on behavior, institutions and behavioral conventions, interdependence, and finally, shared stewardship of shared resources'.

Several investigators have reported finding compelling evidence of vibrant online communities. Rhinegold's [2] classic book about the emergence and growth of The Well provided an extraordinary view into one of the early and more interesting online communities. A more recent study of Internet newsgroups found that a significant number of participants report a sense of community [3]. Furthermore, an individual's sense of community appears to be directly related to the effort and time devoted to the community conversation. A similar study of newsgroups found evidence of online community based on the significant number of personal relationships that were developed and evolved over time [4]. Evidence of community formation online has also been presented for web-based discussion forums [5] and email discussion lists [6, 7].

Online communities vary along many dimensions, including size of membership, general interest and social norms. One fundamental distinction that can be made is whether the online community is geographically disbursed or relatively concentrated. For example, many online communities are built around or grounded in a physical community. Examples include small or rural towns [8, 9], university towns [10], or residents from common school districts [11]. While geographically distant members often join these online communities, much of the interaction is grounded in many of the aspects of the physical place. For example, in one of these communities, the discussion included local news, community calendars and announcements, and local business information [10].

A second type of online community has been built upon relationships between individuals who are geographically dispersed. Some of these communities are created out of personal interests such as a shared medical condition [7]. Others are formed out of work or professional interests such as health care [12], engineering [13], or medical and scientific research [14, 15]. In many cases, these groups are comprised of members who have never met each other face to face, but have nevertheless developed strong friendships and emotional ties. In other cases, the online communities are also participants in the face to face meetings and gatherings of their respective professional organizations.

In this paper we will discuss an investigation into an online community of geographically distributed news reporters. We became interested in this group as a result of an earlier project in which we investigated early adopters of Internet technology [16]. In that study, we used a variety of field research methods, including on-site visits, electronic and phone interviews, and web foraging to understand the many ways that news reporters use the Internet. We learned that one group of technology leaders, still pioneers within their own newsrooms, had created an electronic community of like-minded reporters.

This online community of journalists is arguably different from many other online groups in that it was specifically created to support an already established professional organization. Clearly, this professional community would exist with or without an electronic discussion group. Nevertheless, our initial observations about the discussion prompted us to consider several questions. Specifically, we were interested in what information was important to this community, and in what ways did this online group support the creation and sharing of information among its members? We were also interested in understanding the commitment and the extent of participation of individual group members.

General method

As was mentioned above, our interest and understanding of reporters' use of online technology developed over a period of time and was shaped by and grounded in our earlier research. Our understanding of the online community of news reporters that is presented here was based on participant observation of the online conversation within this community. We joined the discussion group at the invitation of a news reporter and our active observation of the electronic discussion spanned about six months from November 1996 through May 1997. This particular discussion group is not moderated and participation is freely available to the public.

It should be noted that while there are many kinds of electronic discussion groups found on the Internet, the group that is described here is supported by the simple and widely used LISTSERV software. The software supports a group mail list and can be either public or private, and either moderated or not. Over 50 such discussion groups were recently identified in an index of communications-related lists.

A searchable archive of the electronic discussion is also available on the Web. This allowed us to examine archival files of the community discussion for the preceding two years. Altogether, we were able to examine over 4300 messages that spanned a period of 34 months. The message summary information that is typically available in the message headers, such as sender name, time, date, and subject were analyzed using several quantitative analysis tools. A content analysis of the messages was completed by examining a sample of approximately 1800 messages using two generally available qualitative analysis tools. These tools allowed us to index, search, code, and subset the conversation corpus.

General characteristics of the group

There were many challenges inherent in trying to visualize the activity of a large online community. This was especially true when trying to understand changes in the community activity over a significant period of time.

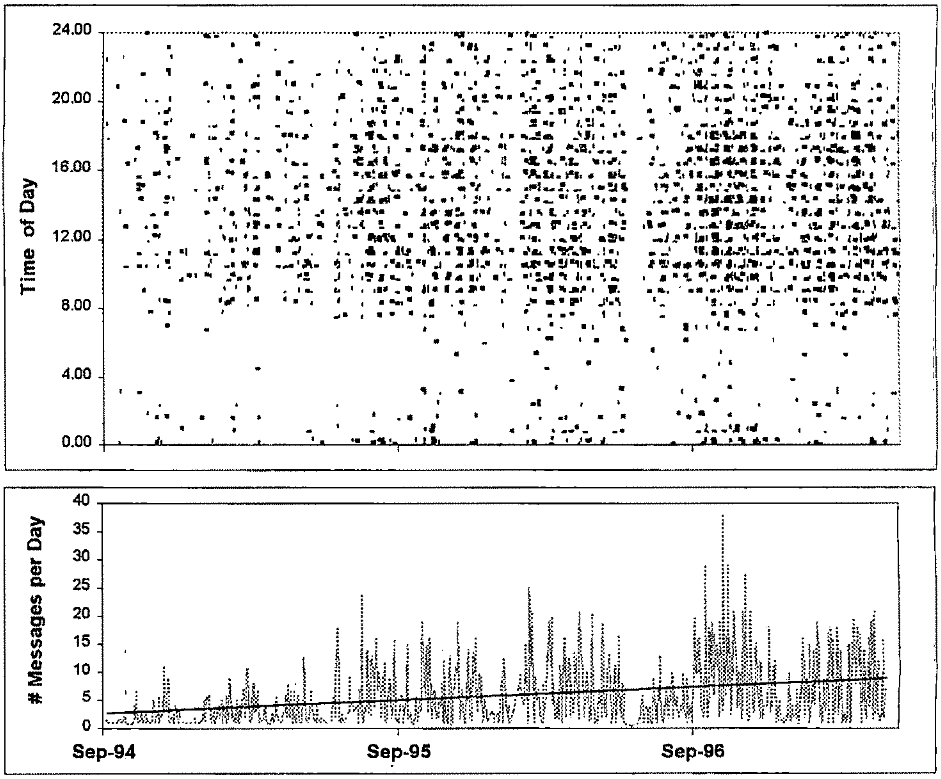
We found that a 'community graph', which can be seen in Figure 1, was a useful way to see the general activity for an online community. The top panel of the display reveals a scatter plot of all message submissions as a function of the time and the day that the message was sent. The bottom panel shows the total number of messages per day and a simple linear trend line.

An analysis of the message activity revealed that there was a general increase in message activity over time. This can be seen in the increasing density of postings in the top panel, and in the increasing daily totals in the bottom panel. The linear trend line also revealed a moderately positive slope. The daily number of messages posted was between 10–20 messages by the end of the analysis period. The message postings in the top panel also showed that this community was largely active during traditional business hours with some activity in the evening.

As can be seen in Figure 1, there was a substantial dip in message activity in the summer of 1996. The group discussion suggests that this was due to problems that resulted from a change in the discussion list software. While not part of this study, it would be of interest to understand how such a technology-based disruption affects community membership and longevity.

A slightly different view of the temporal message activity is presented in Figure 2. The daily message activity was consistent with a traditional office work schedule, i.e. with a similar pattern of messages from Monday through Friday and a significant decrease in message activity on the weekend (Saturday

Figure 1. *Online community activity between September 1997 and May 1997.*



and Sunday). The pattern of activity from Monday through Friday also revealed a rise during office hours starting at about 8.00 am, a peak between 10.00–11.00 am, and then a tapering off of activity after about 6.00 pm. Of little surprise was the near zero activity in the middle of the night between the hours of 2.00–5.00 am. The weekend message activity was lower and more evenly distributed during the day. The differences that we observed between weekday and weekend message activity could be due to several factors. One explanation is that this group of workers followed a social norm to curtail work during the weekend. Another explanation is that this group does not yet have access to the Internet on the weekend.

The personal signatures at the end of posted messages revealed that this discussion group was geographically dispersed and largely based in large cities within the United States. A cursory examination of names revealed that there were many more male than female active contributors to the list.

Commitment and generalized reciprocity

Sharing knowledge with other online community members is the *raison d'être* of this discussion group. The commitment to helping others was visible again and again. A technical question almost always had multiple helpful responses. In some cases, there were multiple responses from the same person as he or she continued to think about a particular problem. Corrections and clarifications were also offered as appropriate. Critical advice about an approach to a news story often revealed a genuine concern that the reporter had thought critically about all of the important issues.

In several cases, care was taken to make sure that other, particularly new, community members felt comfortable asking for help. The often heard phrase 'no question is a dumb question' was posted by one participant and appeared to be the general rule. We observed students asking for a professional critique of their work, and professional trainers asking for a critical audience.

There were, however, widespread differences in the degree of participation in the public discourse. An analysis of the number of messages posted per contributor suggests that there are several different types of list members. As can be seen in Figure 3, which shows the cumulative percentage of messages posted by a cumulative percentage of posters, a small percentage of the list contributors (~15%) account for about 50% of the mes-

Figure 2. *Message activity as a function of time of day.*

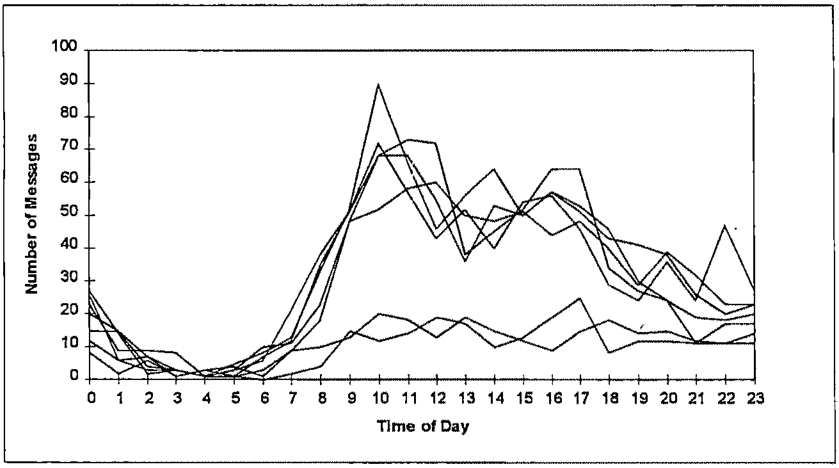
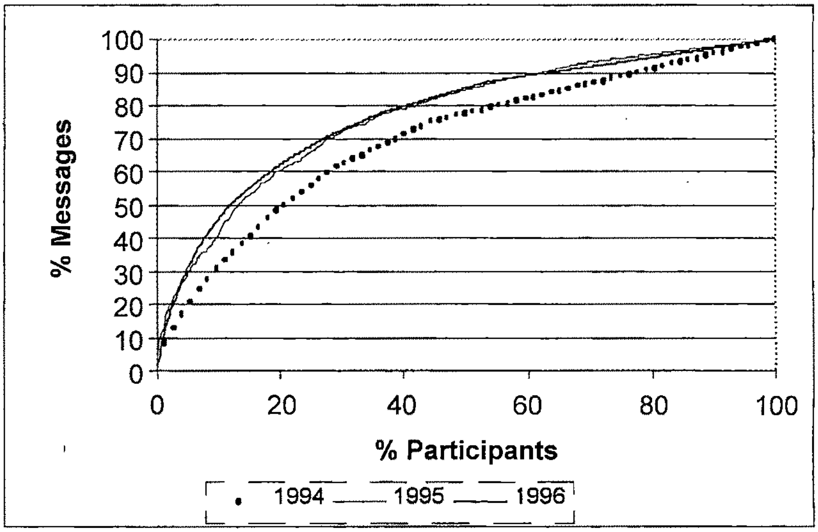


Figure 3. *Cumulative message contributions by community members.*



sages. This percentage is remarkably consistent across the entire sample of messages.

Additional analysis of the messages also revealed that a second group contributed sporadically, or in small bursts, when a topic of personal interest was discussed. Occasionally, news professionals, who were not wired or not participants of this list, contributed to the conversation by asking a member to post a message on their behalf. A third group contributed very infrequently, while an unobservable fourth group presumably 'lurked'.

There is evidence that this electronic community is healthy and can reasonably be expected to continue. The discussion group was formed in September 1994 and there were 72 active participants during the first four months.[1] The number of contributors to the discussion has grown each year and there were ~500 contributors to the list in 1996. As was mentioned above, there has been an increasing number of messages each year as well.

If we label the 1994 contributors as the 'founding members' of the list, then of the twenty most active founding members, ten remained active contributors in 1997. Four of the most active 1995 contributors were also among the most active in the first half of 1997. It would be interesting to understand whether previously active participants have left the community altogether or are simply lurking. The top twenty contributors to the discussion group over the analysis period are plotted in Figure 4. Almost everyone has been a member of the electronic group from the beginning and continues to contribute at significant levels. Clearly, this is a group that will continue to freely discuss professional issues together for the foreseeable future.

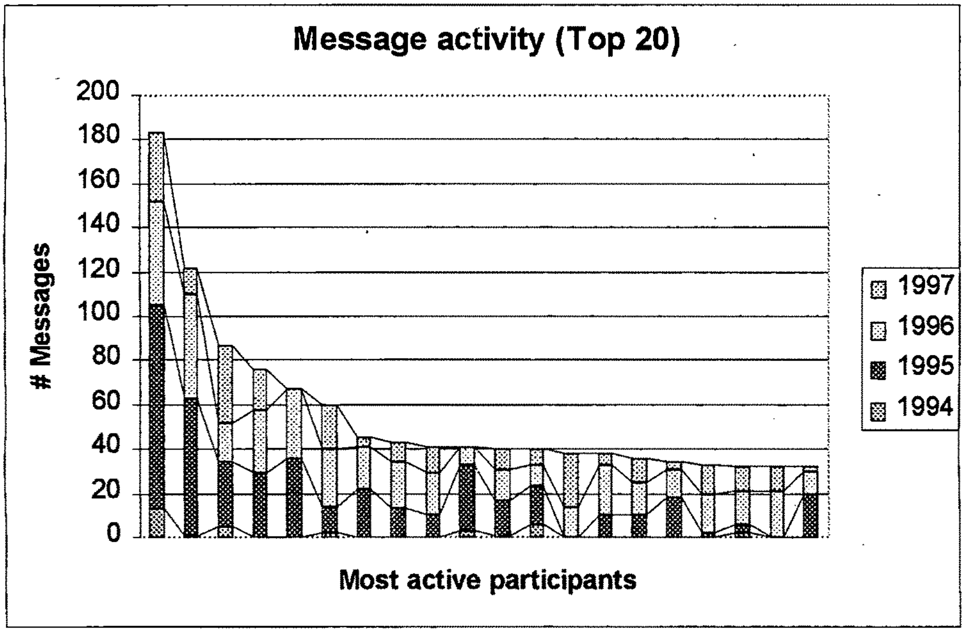
Creating and sharing collective goods

It has been argued that an essential characteristic of an online community is the creation

---

[1] The number of individual participants can generally be obtained from the visible message header information. The total number of subscribers (recipients) to the list is not easily obtained. There is also no way to know whether individual subscribers have actually read the list messages.

Figure 4. *Total messages posted by active members.*



and management of collective goods [2, 5]. This online community of reporters arguably creates two kinds of collective goods:

1. research methods or practices for computer-assisted reporting;
2. informal referrals (and hyperlinks) to important data, news sources, or experts.

These goods are created informally, over time, and are embodied in the electronic conversations of the group.

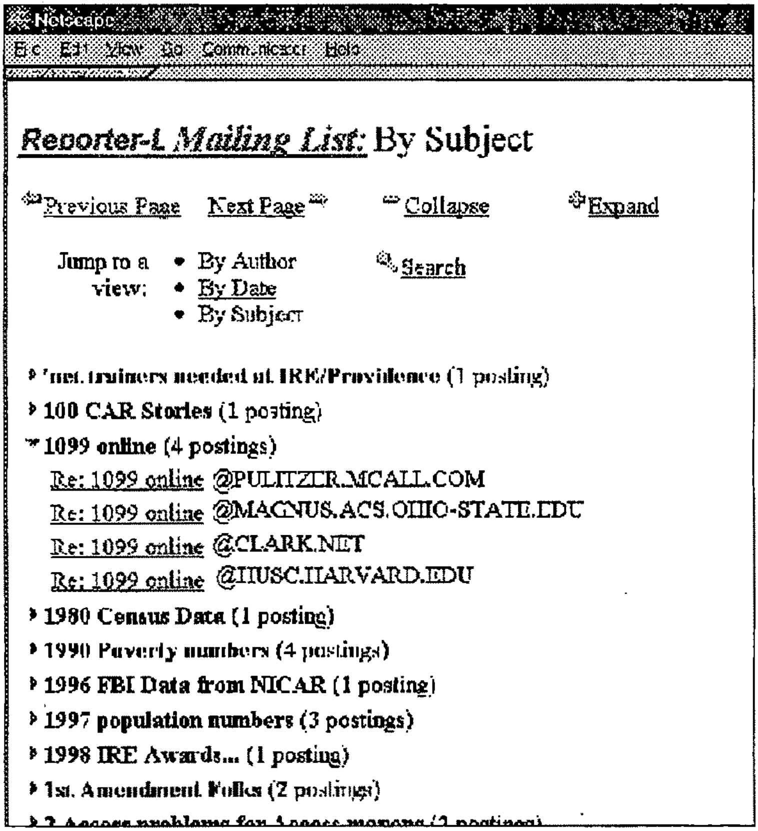
To better understand the content of the conversation in the group, we analyzed a small sample of the list activity (1–15 November 1996). We assigned each message to one of the primary categories listed in Table 1. We found that over 50 per cent of the messages were answers to technical questions. Interestingly, the original technical questions only accounted for 12 per cent of the traffic. Multiple responses were offered for each of the questions and message threads often exceeded 20 messages.

The content analysis of the online conversation revealed that the most frequent contributions are questions and answers to technical problems. Examples included someone asking for help with a FoxPro database program, and someone else seeking help with an analysis of a government spreadsheet. It was not unusual for the answer to include sample source code or example database scripts. Occasionally, there were references to earlier discussions about similar problems. The answers to technical questions

Table 1. Type of messages

Type of posting to Listserv	**%**
Technical questions	12%
Technical answers	51%
Request for information sources, ideas, etc.	7%
Reply with information sources, ideas, etc.	8%
Job announcements	5%
General announcements	10%
Professional group business	4%
Personal and other	3%
Total	**100%**

Figure 5. *Example of archive user interface.*



Reporter-L Mailing List: By Subject

Jump to a view:
- By Author
- By Date
- By Subject

- 'net. trainers needed at IRE/Providence (1 posting)
- 100 CAR Stories (1 posting)
- 1099 online (4 postings)
  - Re: 1099 online @PULITZER.MCALL.COM
  - Re: 1099 online @MAGNUS.ACS.OHIO-STATE.EDU
  - Re: 1099 online @CLARK.NET
  - Re: 1099 online @HUSC.HARVARD.EDU
- 1980 Census Data (1 posting)
- 1990 Poverty numbers (4 postings)
- 1996 FBI Data from NICAR (1 posting)
- 1997 population numbers (3 postings)
- 1998 IRE Awards... (1 posting)
- 1st. Amendment Folks (2 postings)

included both specific answers (e.g., script code) as well as general advice about the way to approach a problem. Members of the group generously contributed these research practices.

In addition to asking technical questions and providing practical, actionable answers, the online group also asked for help in finding information sources. For example, there were often queries asking about the availability of government data sources such as census data, court filings, and tax data. The discussion group also informed community members about the availability of the information that had been aggregated by the ongoing activities of the original professional organization. For example, the professional organization's collection of newsletters, archived news stories, training materials, and shareable data sets are cataloged and the means for member access is described.

The formal management of the online discussion archive is handled by a professional news reporters' organization. A multi-year archive of the complete online discussion is available on the Internet. A sample screen shot of the archive interface can be found in Figure 5. This archive is browsable by date, by thread (i.e. topic of conversation), and by person. In addition, the contents of the conversation archive are searchable by keyword.

In practice, this easily accessible and searchable archive of the group interaction is the community's most important collective good. Technical problems that have been discussed in the past can be easily retrieved and pointers to additional information resources remain accessible. Further, the discussion archive serves as an informal record of community 'experts'. Knowledgeable participants can usually be identified in the online discourse and follow-up private information exchanges can be initiated. Future enhancements to the archive could include social filtering techniques (see, for example, Terveen, Hill, Amento, McDonald and Creter [17]), which could help to identify community experts and to facilitate the dissemination of community knowledge.

Summary

The online community of journalists that we studied was founded on a commitment to helping others and collecting and sharing information. The conversation could be characterized as both concentrated as well as co-operative. There is a general feeling of openness in asking for assistance and providing answers, although a small number of participants account for a majority of the online conversation. This community provides a valuable service to its members by announcing shared information resources, and creates a valuable collective good in the archive of the public discussion.

References

1. Dray, S.M. and Siegel, D.A. Technology and the diffusion of community. *Proceedings of the 13th Triennial Congress of the International Ergonomics Association*. 1997.

2. Rheingold, H. *The virtual community: homesteading on the electronic frontier*. Reading, MA: Addison-Wesley, 1993.

3. Roberts, T.L. Are newsgroups virtual communities? *Proceedings of CHI'98: Conference on Human Factors in Computing Systems*. New York: ACM, 1998.

4. Parks, M.R. Making friends in cyberspace. *Journal of Communication*, 46(1), 1996, 80–97.

5. Erickson, T. Social interaction on the Net: virtual community as participatory genre. *Proceedings of the Thirtieth Hawaii International Conference on System Sciences*. 1997.

6. Millen, D.R. *Information technology and the computer-assisted reporter*. Paper presented at CASTAC (Committee on the Anthropology of Science, Technology, and Computing) Conference, Renneselaer Polytechnic Institute, Troy, New York, 1997.

7. Preece, J. Empathic communities: reaching out across the Web. *Interactions*, 5(2), 1998, 32–43.

8. Franzke, M. and McClard, A. Winona gets wired: technical difficulties in the home. *Communications of the ACM*, 39(12), 1996, 64–66.

9. Schuler, D. Community networks: building a new participatory medium. *Communications of the ACM*, 37(1), 1994, 39–51.

10. Carroll, J.M. and Rosson, M.B. Developing the Blacksburg Electronic Village. *Communications of the ACM*, 39(12), 1996, 69–74.

11. Kraut, R., Scherlis, W., Mukhopadhyay, T., Manning, J. and Kiesler, S. The HomeNet field trial of residential Internet services. *Communications of the ACM*, 39(12), 1996, 55–63.

12. Bowers, L. Constructing international professional identity: what psychiatric nurses talk about on the Internet. *International Journal of Nursing Studies*, 34(3), 1997, 208–212.

13. Donnelly, R. and Hermann, R. Building an Internet resource for a specialized online community. *Proceedings of SIGDOC '94: technical communications at the great divide*. New York: ACM, 1994, 190–197.

14. Schoch, N.A. and Shooshan, S.E. Communication on a Listserv for health information professionals: uses and users of MEDLIB-L. *Bulletin Medical Librarian Association*, 85(1), 1997, 23–32.

15. Walsh, J.P. and Bayma, T. Computer networks and scientific work. *Social Studies of Science*, 25, 1996, 661–703.

16. Millen, D.R. and Dray, S.M. Job transformation in the age of the Net. *Interactions*, 49(2), 1997, 13–18.

17. Terveen, L., Hill, W., Amento, B., McDonald, D. and Creter, J. PHOAKS: a system for sharing recommendations. *Communications of the ACM*, 40(3), 1997, 59–62.

‘Hyped’ text markup language. XML and the future of web markup

Claire Warwick and Elliott Pritchard
Department of Information Studies, University of Sheffield, Sheffield, S10 2TN
c.warwick@sheffield.ac.uk

There is a widespread perception that, in terms of web-technology, XML is going to be the ‘next big thing’. Given the amount of comment that it has generated, it seems to be on its way to achieving that status. But how much of the praise should be taken at face value, and how much of the hype is credible? In the following article we examine some of the claims made about the importance of XML and consider how far the enthusiasm about it can be justified. Will XML cause a revolution that will change the way that everyone uses the Internet, whether as searchers or data creators? Or is it a tool for certain types of e-commerce and large-scale markup, which may not have a significant impact on the majority of web users?

HTML and its discontents

Many XML articles begin with rhetorical questions such as ‘HTML is the most successful document format in history, why would anyone want to go beyond it?’ [1], or ‘Thanks to the abundance of software and literature available anyone can create web pages now. So what’s the point of a new language, and furthermore one as complicated as XML?’ [2]. These are followed by a glowing description of XML intended to clarify exactly what the point is. But these questions are valid ones. In the clamour for added sophistication in web-markup, some commentators seem to have forgotten that for all its faults HTML suits the needs of many people who wish to mount basic web pages.

There can be little doubt that part of the success of the World Wide Web is due to HyperText Markup Language (HTML), a simple means to format data for presentation in a web browser. Tim Berners-Lee, the creator of the Web, fashioned it as a specific application of a pre-existing standard for document markup, the Standard Generalized Markup Language (SGML). The winning characteristic of HTML is its simplicity. The Web may not have achieved such widespread use if so many people had not found HTML so easy to learn and to write. However, it is this very simplicity that is now cited as HTML’s main drawback.

The success and the limitations of HTML stem from the same cause. HTML is a markup language that describes the way that a page looks. This is its great strength since it is easy to learn and use. However, it also proved a weakness, in that a small number of tags limited what could be done with it. One problem with HTML, which is simultaneously a reason for its success, is error-handling. Netscape designed browsers so that they would read broken code, and be forgiving of writers who forgot to close tags. This was important in the early days of the Web, when few WYSIWYG editors existed, and most markup was done without the benefit of a parser to check its validity. It allowed non-experts to mount pages, which would then be accessible, even if invalid [3].

Error handling also means that browsers will interpret HTML tags in different ways, which has led to a divided and chaotic web. In an attempt to control the page some designers specify which version of a browser and what screen resolution should be used [4]. If this is a high specification the page may be inaccessible, load slowly, or look different from the way the designer envisaged on older machines. This leads to various problems. For those who use the Web as a way of accessing information content, this may prove frustrating, if they cannot load the page.

Corporate clients who depend on a certain style for all their products may also be unhappy with their inability to regulate the look of their web site. Such efforts also mean that the site must be overhauled as technology changes. If a new version of a browser appears, designers may be faced with the task of redesigning thousands of pages in a site to suit the new specification [5]. For the site provider this is likely to be costly.

The forgiving nature of HTML allowed browser vendors to build on the original specification by introducing their own tags, once they realised that HTML was too simple for their needs. The job of altering the specification should have been that of the World Wide Web Consortium (W3C). Instead it failed to keep pace with new features that companies were implementing [5]. This timelag is explicable when we look at the nature of W3C itself. The W3C is as close to a governing body of the Web as it is possible to have. Its head is Tim Berners-Lee, and the importance of it may be judged by the list of its member organisations, which include Microsoft, Netscape, Sun Microsystems, IBM, and Adobe. Any company can join. On payment of a fee they can take part in the creation of new standards. The remarkable feat of the W3C is that competing organisations such as Microsoft and Netscape work together to bring about industry-wide progress [6]. A successful idea from a W3C working group becomes a working draft, which is then submitted to the director. If he declares it a 'Proposed Recommendation' it is forwarded to all member-organisations, whose votes determine its progress. The idea may become an Official Recommendation, a working draft again, or be abandoned. In practice this process, and the refinement of a document, can take a great deal of time. It is therefore not surprising that it failed to keep pace with the changes in HTML.

The proprietary extension of HTML brought about more problems than it did solutions and it became clear that HTML was not capable of taking the strain. The W3C wanted to retain the essential simplicity of HTML, but at the same time to meet the growing demands of the user community. Different elements of the community argued for different developments, according to their needs [7]. It had become apparent that the problem was inherent in HTML itself. HTML could not be turned into the advanced, flexible tool that some required, simply by adding elements to it. This would only compromise the simplicity that made HTML so successful [8].

The solution arrived at was to produce a markup language which was extensible, which could be extended to fit a user's requirements by allowing them to create their own tags. XML is more than just HTML with add-ons, however. It is also a different concept of markup. Whereas HTML tags describe the appearance of a web page, XML tags describe the information [9]. When XML was first developed, some thought that it would be a straightforward replacement for HTML [10]. Since then, it has become apparent that this is not the case [11]. HTML is already too ubiquitous for this to happen, and it is also the appropriate tool for many users and providers [12]. The vast amount of information already coded in HTML is enough to guarantee its persistence. For those who wish to perform simple web authoring, it is also the preferred language [13]. Bodies who regulate standards have accepted this, since the ISO standardisation of HTML implies a conviction that HTML will persist for at least 25 years. W3C has also reformulated HTML as an application of XML, which recognises that the simplicity of HTML is attractive to many web users and smaller information providers [14]. The situation for large corporate web sites and information providers, however, might be different. As Flynn argues:

> HTML is at the limit of its usefulness as a way of describing information, and while it will continue to play an important role for the content it currently represents, many new applications require a more robust and flexible infrastructure [15].

The genesis of XML

The language of many articles about XML is full of almost religious fervour about its potential. But before joining the community of the converted it is important to be aware of the development and status of XML. As we showed above, the history of HTML dictated its use, and its advantages became its drawbacks. To examine whether the same is true of XML we must look at how it developed.

The W3C's SGML Working Group, formed in August 1996, was responsible for the conception of XML. The first XML Working Draft was announced at the SGML '96 Conference in November 1996. Successive versions of this draft were completed in 1997, and finally a Proposed Recommendation was announced in December at the SGML/XML '97 Conference. Voting by member-organisations took place, and on 10 February 1998 XML 1.0 was announced as a W3C Recommendation. As the '1.0' suggests, successive versions of this specification may appear. Current XML-related working drafts include eXtensible Stylesheet Language (XSL), XML Linking Language (XLink), and XML Pointer Language (XPointer).

XML, like HTML, was developed from SGML (Standard Generalized Markup Language). It is therefore not surprising that when users became disillusioned with HTML, they looked again at SGML. This language, which is best described as a way to define and enact specific 'document types' that describe the structure and content of electronic documents, was developed from work done at IBM in the sixties. In 1978 the American National Standards Institute (ANSI) created its first version of SGML as an attempt to standardise the way of defining and using markup in documents and it became an ISO in 1986.

Since there is already a markup language which is more complex than HTML and which describes the content as well as the look of a document, why was it necessary to create yet another subset of the existing language? Why not simply use SGML on the Web? The main reason for this is the size and complexity of SGML. If HTML was too limited, then SGML is far too complex for the needs of web users. When W3C created XML as a subset of SGML they omitted complicated or unnecessary parts in order to make it usable over the Web [16]. The scale of the simplification may be judged by the fact that the formal definition of XML takes up 33 printed pages, as opposed to the approximately 500 pages used to define SGML.

It was important to retain the link with SGML rather than creating an entirely new language. Many large commercial organisations had large amounts of data in SGML, which they had amassed in the previous ten years. Clearly they did not wish to incur the costs of converting it to HTML or to an entirely new web language [17]. Because of the heritage of XML, companies with SGML now find that their information is much more 'web-compatible' than in the past [18].

Not only has XML meant a shrewd simplification of SGML for the Web, it has meant a shrewd re-naming. Tim Bray, one of the editors of the XML 1.0 Specification says:

> There were several acronyms that we considered. I believe there was MGML, for Minimal Generalized Markup Language, and something called SIMPL for Simple Internet Markup Protocol, or something like that. Eventually we voted, and XML – for Extensible Markup Language – won out. It was short and sweet, and people liked it [19].

The name may appear inconsequential, but it plays an important part in marketing this new standard. Judge and Ogg suggest that:

> Where SGML is a 'square' thing for document handlers, XML is a 'hip' thing for web monkeys. The XML people, most of whom are old SGML hands, are blinking to discover that third year computer science students are writing XML products – SGML never had that kind of street cred [3]

See also Kimber [20].

As a marketing strategy, it has worked well. XML now has the advantage of apparent novelty and has brought about a fresh wave of excitement about web markup. As Zelnick puts it:

> Appropriating XML is a way to get the attention of reporters deluged with press releases...Why do SGML vendors call themselves XML companies today? [21]

In the world of the Web, especially that of e-commerce, the new is always assumed to be better. XML could be regarded as SGML repackaged; this is ideal for companies who need something new to promote. However, it is confusing for users. They are fed a diet of hype, but given few facts with which to judge how important XML might be for them.

It is not surprising, thanks to the successful publicity generated by the advent of XML, that it is not difficult to find products associated with XML. As Freter puts it:

> the world has responded to XML. Go to any trade show or conference associated with publishing, documents, or the Internet (and intranets or extranets), and you will see vendor upon vendor pledging or even demonstrating support for XML [22].

The Web is full of such pledges of allegiance, or announcements about new XML parsers, or editors, or projects. Yet it is still relatively difficult to web sites that actually use XML.

Why should this be the case, if XML is really the answer to so many problems? Part of the problem is that associated specifications such as XSL, XLink, and XPointer have yet to achieve recommendation-status. These are the languages that govern how pages should look. XSL, (Extensible Style Sheet), XLink and Xpointer are languages that govern linking within XML documents and to other documents. Yet changes in these are unlikely to be major, and this should not delay the introduction of XML.

More significant is that the number of resources available to help one learn XML is still few. If XML is to achieve widespread use then such resources are needed. The official specification, although short, does not make a good introduction to the subject. Its technical phraseology, although appropriate for a formal definition, is not geared towards a beginner [23]. Although there are numerous online resources to teach HTML, there are relatively few for XML. As Flynn comments in the XML FAQ, 'there is still a need for tutorials, simple tools, and more examples of XML documents' [15].

We also lack a web browser that supports XML fully. Version 5.0 of Microsoft Internet Explorer offers partial support, but both Microsoft and their rival companies have pledged full XML support in the next version of their browsers [24]. Microsoft, whose XML implementation-efforts began sooner than their competitors, have plans to integrate XML into Office, as well as their next browser [25]. This is important news and Hogan argues that this virtually assures the broad adoption of XML [26]. However, until we can use an actual, as opposed to promised, web browser it is impossible to find the few sites that use XML or to make the most of added XML functionality. This in turn inhibits developers from using it, as they are unwilling to spend time and money converting from HTML. This becomes a vicious circle of potential but lack of use.

Although some people have begun to become suspicious as to whether XML will ever be implemented, Sharpe argues that:

> when a technology has the blessing of the W3C and is endorsed by hundreds of industry participants, including such formidable players as Microsoft and Oracle, you can be sure it is here to stay [27].

His confidence may seem overstated. However, so many vested interests have a stake in the success of XML that every effort will be made to assure its adoption. However, its success is not necessarily inevitable. Even Bill Gates has been known to be wrong in the past, notably about how popular the Internet might prove.

A question that we must ask then is: since the success of XML seems likely, what will the scale of that success be? It seems likely that some people are going to implement XML. But it is still unclear how widespread its adoption will be, and how quickly it will happen.

The XML revolution

Some consider that 'XML is the future of the Web' [28]. It will, they insist, instigate the Web's 'second-coming' and be 'as big as the Web itself' [29]. Such enthusiasm may become self-fulfilling. The more hype that is generated, the more people will hear of XML, the more XML will be implemented. Wild promises are being made, such as:

> The impact of this revolution will be as dramatic as the creation of the Internet itself. Information overload will give way to dynamic, targeted, granular and personal information manipulation. The Internet will be transformed from a massive unstructured, unmanaged and cumbersome collection of documents, into a structured, interactive, navigable and useful part of our everyday lives [26].

The above theory seems to presuppose that once XML comes into use, everyone will use it. An entire web coded in XML would certainly be dramatically different from the current web, but the Web will not transform itself overnight. Those prepared to put the extra effort into XML coding may reap its benefits. But it seems foolish to expect that everyone will make that extra effort. We may choose to be enthusiastic about XML, however, it is important to retain a sense of balance about its potential impact. It is therefore important to look at the features of XML, and assess whether they are likely to be beneficial or detrimental to those who are considering adopting it.

Complexity

Anyone who is familiar with HTML will not be daunted by XML-code, since it looks very similar. To someone unfamiliar with markup, XML may be even more comprehensible, because unlike HTML, its tags tend to be in natural language. For example, it is unlikely to be apparent what the HTML tag <hr> signifies. However, the use of an XML tag such as <TelephoneNumber> is obvious [30].

XML also allows users to choose their own tag names, and as long as they are intuitive, it is arguable that XML is clearer, and therefore easier to implement, than HTML. But this is not the case. One of the great benefits of HTML is that it is quick and easy to write. XML, however, although not too difficult to learn, is part of a larger puzzle. Someone wishing to write web pages in XML may have to learn the associated specifications XSL, XLink, and XPointer as well. Altogether, this package is neither quick nor easy to master.

XLink and XPointer are extremely complex languages, designed to solve a simple problem. In HTML it is possible only to make one way links, that is from your page to another. This may confuse users, who may follow several links and then find it hard to find their way back to the original document. This is not desirable if they start from a commercial site that seeks to keep its customers from wandering off before they make a purchase. SGML, however, allows two way links, which enable the user to return to the original document after they have followed an outward link. This is a relatively simple piece of markup in SGML, however, for reasons that remain obscure to all but XML specialists, this simple operation has necessitated the writing of two entire specifications in XML. This seems a mistake on the part of XML developers. The complexity of the additional specifications may prove a disincentive to their use by all but the most specialist users. Many information providers who might have liked to use two-way linking in basic HTML documents may be daunted by this heavy handed approach to the additional specifications. Bosak therefore feels that full XML will, at least initially:

> be the domain of a few experts working on a large, specialized publishing application by hand. These will be the applications that demand the highest level of automation and media independence – newspapers, business directories, encyclopaedias, commercial catalogs, television schedules, and so on [31].

If the extra sophistication that XML brings to web markup also brings added complexity, it may hinder its widespread use. We can foresee a scenario in which small information producers are likely to persevere with HTML, while large commercial concerns hire technical experts to deal with huge amounts of XML. Users may therefore be faced with the same chaotic, divided web that XML enthusiasts wish to avoid.

Separating style and content

Whereas HTML tags describe the appearance of information, XML tags describe its meaning. Appearance is as crucial as ever though, and this is the reason for the stylesheet language XSL (eXtensible Stylesheet Language). XSL tells a browser how to present the information found in an XML file. But because it is not part of the XML file, style and content have been separated.

This is important when the style needs to stay the same and the content change [32]. XML enthusiasts argue that this feature offers a number of benefits. For example, a stylesheet could be used as a corporate document-template. Every time a new document is added to the company web site, it could use the same stylesheet and therefore conform to the same format. Stylesheets might also be used off-line or on-demand to create personalised web sites in response to user requests [27].

XML is also being used to create a version of the Web for use on a small screen or a cellular telephone. It might allow simple navigation or even convert text into speech [33]. Stylesheets mean that the same web site can be viewed in the usual way via a PC or on the phone without developers spending time creating two separate web sites. They could also be used to aid accessibility for visually impaired users. Current browsers of this type are unable to deal with the use of frames. Separate stylesheets would mean that content was still accessible, whatever the design [30].

Although these benefits are important, it is already possible to use the Cascading Style Sheets (CSS) language in conjunction with HTML. This would achieve the same results without the use of XML. It also has the benefit of being associated with an already familiar language of relative simplicity. It is, however, relatively rare, and it may be that XML (which requires the use of a stylesheet) will be important in encouraging the separation of style and content on the Web. It is equally possible, however, that if the use of CSS and HTML becomes popular, there may be a feeling that the effort of learning and converting recourses to XML is unnecessary.

Error-handling

Connolly *et al* claim that 'we can think of XML as HTML without the 'training wheels' ... the key difference between XML and HTML is that XML will be written properly' [34]. XML, unlike HTML, is not forgiving of errors. This means that anyone who uses XML will have to write valid, or at least well-formed, documents for a browser to be able to display them. This suits the experienced web developer, concerned with the integrity of documents, but its implications for the general web user, who either writes or reads HTML, are less attractive.

The insistence that XML should handle errors strictly came from browser manufacturers [3]. This is understandable, since the extra programming involved in handling errors in HTML code has contributed significantly to the unwieldy file-size of web browsers. Error-handling is not something that the browser-vendors want to encourage, since strictness saves them a great deal of time and effort [35].

Web page creators are less likely to welcome this, since a single error in code will prevent the entire document from being read [33]. That should not cause problems for professional web designers who are paid for correct markup. But a great strength of HTML was that anyone could use it. If they forgot to close one of their tags, the page would still be displayed. This attractiveness does not simply apply to amateurs building a home page, but to any organisation without the means to spend large amounts of money on designing and maintaining a web page; for example small firms, charitable and voluntary organisations, and many libraries and academic departments. The strictness of XML error-handling is not user-friendly. Many people who maintain web pages in such organisations perform this role as part of their overall job. If this sort of busy information professional is faced with the choice of forgiving HTML and difficult XML, there will be little incentive to move from one to the other. They are unlikely to have the time or the inclination to master such an unfriendly language. This ought to worry those with a commercial interest in XML. If it is not widely used, they may find that the market for this streamlined software, made possible by strictness over errors, will be smaller than they had imagined.

Context

Any web user knows that searching can be frustrating. This is often because the search terms used lack context [36]. Metadata can alleviate this problem, as Rath explains:

> The search in Alta Vista for an SGML book written by Mr. Pepper brings more than 200 hits and none of the first twenty hits are useful...[a] search at amazon.com lists the book immediately, but here the search is done in a database with metadata and not in HTML pages [37].

Metadata can prove very useful as a way of adding value to raw data. XML handles metadata well, since tags describe the information that they contain not simply how it should be presented [38]. It therefore provides a way to incorporate metadata into the body of a file, and can provide context for information in the document. As Hogan argues:

> 1000 might be a good price, in dollars for a state-of-the-art laptop. However, it would be a very bad number of days required for delivery. As a result, the tag that puts the 1000 in context is critical [26].

XML therefore allows users to compare like with like, and control the context of their searches [39].

The example of a search of Amazon.com is particularly pertinent, as it is in the area of e-commerce that XML metadata is most likely to be used. A book buyer might wish to compare prices available on various sites. A search engine could search for a title by a specific author and tell you where you could find the cheapest price. This is not only good for the customer, but for the bookseller, whose bargain is easier to find [40].

This kind of searching would be far more sophisticated, and more successful, than what we have today on the Web. HTML allows for the incorporation of metadata, using the <meta> tag, which goes in the header of the document. However, it is open to abuse. The addition of common keywords unrelated to the topic of the site has been used as a way of ensuring that sites appear high on search engine ranking lists. This technique has lead to some search engine bots deliberately discounting meta tags. Because XML tags describe the content of the page this is much harder to abuse. If the laptop maker's page uses a tag called <price> and I search for 1000 within that tag, it will not be found if the tag name has been changed to incorporate irrelevant, but popular terminology. This does not preclude abusers from using their own tags, to allow those searching to find what they are looking for. But it is less likely that unwary users will come upon material that they find offensive when entering seemingly innocuous terms.

The use of specific XML tags to describe information sounds full of promise. However, like the language that it is supposed to be outstripping, one of XML's strengths may prove to be its greatest weakness. XML allows the user to make up the names of tags they wish to use. Unlike HTML, where, for example, <p> must be used to indicate a paragraph beginning, in XML a user who wanted to markup prices could choose to use <price>, they could also use <cost> or <amount>. A search on <price> would not find information contained in the other two tags. This may not apply to popular generic tags, but as the complexity of markup increases it is likely to become more problematic. It is also possible to specify the use of a tag through attributes in XML, so for example a travel company might use <time: type=arrival> to indicate what time flights arrived. Another might choose to nest another tag within time and use <time><arrival>15.30</arrival></time>. As long as these were properly described in the DTD, both would be equally correct, but might not be easily found by the same search.

These problems might be avoided by the development of agreed specifications across industries or communities of interest, who would agree to use a common set of tags. This would be achieved by the creation of agreed DTDs. DTDs (Document Type Definitions) are used to define the names and the structure of the tags in an XML document. They therefore standardise the way information is classified. There may be incentives to develop common DTDs for the kind of reason outlined by Shimmin:

> Consider what would happen if two DTD vendors to the medical industry built applications based upon overlapping yet inconsistent DTDs. Many hospitals would not be able to exchange patient records unless each vendor provides translation applications to parse one another's DTD – neither is a feasible solution [41].

Libraries, museums and academic or voluntary communities of interest might also join together to develop the same kind of interoperability. Communities of interest have a strong incentive to reach such agreement [42]. However, the work involved in doing this should not be underestimated. It is likely to prove time consuming to produce the specification, and this presupposes that all members of the community can reach agreement, or that they would want to.

Particularly in the commercial sector, there might be resistance to the idea of making the parallel searching of the sites of several vendors easier to carry out. The searchability of XML might lead to an increase in sales [43], but surely that would only be for whoever was selling the product most cheaply. Some companies might lose out by coding their sites in XML and adhering to an industry-wide DTD, since companies prefer to keep customers on their own web sites, not to make it easy for them to go elsewhere [44]. Competition could also make it difficult to agree who should create DTDs, and how they might go about the potentially difficult process of persuading their competitors to use them [45]. Even if it were well-constructed, or perhaps particularly if it were well-constructed, agreeing to use a competitor's DTD could be seen as losing face, or, more crucially, commercial advantage.

The area of document definition, then, is another area where XML may face problems. Bremser warns that 'until industries and organizations agree on DTDs, XML will be a localised rather than a universal tool' [46]. The World Wide Web Consortium, have unhelpfully said nothing about standardising DTDs. The unfortunate truth is that 'XML 1.0 is an industry standard, but the DTDs upon which it relies can be as proprietary as the formula for Coca-Cola' [41].

Open standard

XML is an open standard. This means that unlike a proprietary format such as the Microsoft Word Document, XML is not designed for the use of one particular company. The Word document has been designed to be opened, edited, and saved by Microsoft Word. Other word-processing software can convert to and from this format, but other companies are not free to edit it, as it has been designed and copyrighted for the use of Microsoft product. Any company, however, may build XML tools.

This is extremely important, since it means that its use will not be limited to one browser. Anyone should be able to write a web page in XML, which everyone should be able to view. Its status is remarkable, considering the competitive nature of W3C's member organisations. As Bosak enthuses:

> the amazing thing about XML is that all of these people and organizations set aside personal and corporate agendas to co-operate in the construction of a genuinely open standard, driven entirely by user needs [31].

This co-operation has been necessary to ensure that XML should not fall foul of the proprietary problems that beset HTML, since this would be costly, and might lead to its eventual failure [47].

Although XML was designed for the Web, its simplicity of design means that it may be used more widely. It makes a good format in which to transfer files between different applications and to interchange information between computer programs [48]. An important reason for this is the increasing realisation of the importance of the Web to general software companies such as Microsoft. They recognise the potential for making information in all application-formats as exchangeable as possible. Although Microsoft already had a method of exchanging data between its different applications, the use of XML for transferring information has the potential to fully integrate the Web into their Office package, as if it was just another Microsoft application [49].

The XML open standard is intended to level out the playing field so that a company of any size is able to compete against a company like Microsoft. However, there are already doubts about whether this is something that Microsoft will embrace, especially given their well-documented lack of charity to their competitors. Some commentators fear that the main players like Microsoft and Netscape will be unable to resist the temptation of gaining commercial advantage, by adapting XML to their own proprietary needs [9, 47]. Others feel that such fears are exaggerated, but at this juncture it is impossible to predict what will happen [50].

Although both Microsoft and Netscape have co-operated in creating an open standard, the test of their intentions will be in its application. Commentators, however, remain pessimistic about how likely they are to maintain an 'open' approach. St Laurent argues:

> While the browser vendors have supported the XML development process and have applied XML where it seems most advantageous to them, a full implementation of XML that encourages its ubiquity may not be in their best interest.

He also observes that any format that allows free movement between different proprietary operating systems would be a threat to the current business practices of the software industry [51]. Bray also asks 'How is someone with an 80 per cent market of anything going to do the best by their stockholders by supporting open standards? [52]'. It would be understandable if large software companies like Microsoft, having made huge profits using existing business methods, were reluctant to change their ways. It is such doubts that overshadow the possible success of XML as an open standard. Should it become increasingly proprietary, as HTML did, the problems caused by the extension of the earlier language would return to bedevil web users and providers alike.

Conclusion

It is apparent, then, that XML should be treated with caution at present. As with many new developments in information technology, it has been surrounded with hype and ambitious publicity. This has some basis in fact. As we have argued, XML has the potential to be very useful to certain users, notably commercial ones. The facility to separate style from content and the provision of information about that content in the markup represents an improvement on HTML. XML has the potential to make web searching much easier for users and thus to make the dissemination of information more effective for information providers.

However, it has not yet achieved the revolutionary status that some have claimed for it. Lack of browser support means that it is hard to find sites encoded in XML, and may therefore be retarding the development of such sites. The complexity of associated languages, a lack of effective tutorials and strictness in error handling also mean that XML will be a demanding language to learn. This may mean that it is not, at least initially, widely adopted, especially by voluntary or small information providers. The ease of searching that XML promises will also be dependant on the development of shared DTDs and the willingness of the software industry to persevere with a commitment to open standards. There are reasons to doubt the success of these commitments, particularly where it may threaten commercial advantage.

It therefore appears that, at least in the short term, XML will not mean a web revolution. Small information providers may continue to use HTML, trusting its simplicity to provide the features that they need. Large commercial companies may adopt it, and so make SGML data more useful, but others may persevere with the use of large web databases. The lack of standardisation of the vocabulary for XML name spaces may also cause problems. This therefore suggests that the Web is not about to be reborn, nor that it will suddenly become homogenous and easy to interrogate as an information source. XML has potential to have an important impact. It is still too early to judge just how widespread its impact will be.

References

1. Bray, T. Beyond HTML: XML and automated web processing, Netscape. *View Source Magazine*, September 1997. http://developer.netscape.com/viewsource/bray_xml.html

2. Macherius, I. XML: a professional alternative to HTML: expert's revolution. *IX Online*, June 1997. http://www.heise.de/ix/artikel/E/1997/06/106/artikel.html

3. Judge, P. and Ogg, C., eds. If XML does replace HTML, it will be a different kind of web. *WOT Ezine*, 28 November 1997. http://www.wot.co.uk/public/ezines/wot23-281197.html

4. Kennedy, D. The case for XML. *SGML Resource Center*, [no date]. http://www.mcs.net/~dken/xmlcase.html

5. Veen, J. Designing for the future with XML. *Hotwired*, 11 May 1998. http://www.hotwired.com/webmonkey/98/19/index0a.html?tw=xml

6. Lourier, P. and Bray, T. XML – Extensible Markup Language. *Technetcast*. http://www.technetcast.com/hz-show-980424.html

7. World Wide Web Consortium (W3C). *The World Wide Web Consortium issues XML 1.0 as aW3C Recommendation: XML 1.0 fact sheet*, 10 February 1998 – a. http://www.w3.org/Press/1998/XML10-REC-fact

8. Sperberg-McQueen, C.M. What is XML and why should humanists care? *Digital Resources for the Humanities, 1997 conference*. Oxford, 1997. http://users.ox.ac.uk/~drh97/Papers/ Sperberg.html

9. Krantz, M. Keeping tabs online: doing business on the Net is hard because the underlying software is so dumb. XML will fix that. *TIME Online*, 10 November 1997. http://www.pathfinder.com/time/magazine/1997/dom/971110/technology.keeping_tabs_.html

10. Frentzen, J. Will you be ready when/if XML replaces HTML? *PC Week*, 31 December 1997. http://www.zdnet.com/zdnn/content/pcwk/1501/266546.html

11. Singh, P. XML broadens horizons of formatting and presentation. *Internet World*, 14 December 1998. http://www.internetworld.com/print/1998/12/14/intranet/19981214-advisor.html

12. Freter, TXML: it's the future of HTML *Sun Microsystems*, June 1998. http://www.sun.com/980602/xml/

13. Luh, J.C. Lots of hype, but few XML browsers or editors on horizon. *Internet World*, 5 October 1998. http://www.internetworld.com/print/1998/10/05/webdev/19981005-lots.html

14. World Wide Web Consortium (W3C). *XHTML 1.0: the Extensible HyperText Markup Language: a reformulation of HTML 4.0 in XML 1.0*, 5 May 1999. http://www.w3.org/TR/xhtml1/

15. Flynn, P. *et al.* Frequently asked questions about the Extensible Markup Language. *University College Cork*, 6 October 1998. http://www.ucc.ie/xml/

16. ETHOS. The Extensible Markup Language (XML). *European Telematics Horizontal Observatory Service*, [no date]. http://www.sgml.u-net.com/xml.htm

17. Walsh, N. A technical introduction to XML. Author's web site, 18 February 1998. http://nwalsh.com/docs/articles/xml/

18. Bremser, W. Hands-on tutorial bringing XML to life in the real world. *Internet World*, 5 October 1998. http://www.internetworld.com/print/1998/10/05/iwlabs/19981005-xml.html

19. Denison, D.C. The road to XML: adapting SGML to the Web. *XML.com*, 2 October 1997. http://www.xml.com/xml/pub/w3j/s1.discussion.html

20. Kimber, W.E. Why do we need XML? Presented at: *Swedish SGML User's Group XML Seminar*, 8 September 1998. http://info.admin.kth.se/SGML/Konferenser/xml98sve/seminar.html

21. Zelnick, N. Don't buy XML stock – give it time to get real. *Internet World*, 30 March 1998. http://www.internetworld.com/print/1998/03/30/webdev/19980330-stock.html

22. Freter, T. XML: mastering information on the Web. *Sun Mocrosystems*, March 1998. http://www.sun.com/980310/xml/

23. Bray, T. Beyond HTML: XML and automated web processing. *View Source Magazine*, September 1997. http://developer.netscape.com/viewsource/bray_xml.html

24. World Wide Web Consortium (W3C). The World Wide Web Consortium issues XML 1.0 as a W3C Recommendation: XML 1.0 Testimonials. *World Wide Web Consortium*, 10 February 1998 – c. http://www.w3.org/Press/1998/XML10-REC-test

25. Gates, B. Remarks by Bill Gates, Microsoft Corporation, Gartner Symposium, Monday, October 6, 1997, Orlando, FL. *The SGML/XML Web Page*, 6 October 1997. http://www.oasis-open.org/cover/gates-gartnerXML.html

26. Hogan, M. XML: The foundation for the future. *The SGML/XML Web Page.* http://www.

oasisopen.org/html/xml_foundation_future.html

27. Sharpe, B. XML: it's not your father's HTML. *Internet.com*, [no date]. http://www.webreference.com/authoring/languages/xml/intro/

28. STEP Sturtz Electronic Publishing GmbH. Introduction to XML. *The SGML/XML Web Page.* http://www.oasis-open.org/html/step_intro_to_xml.html

29. Eckart W. [product manager for Netscape]. Quoted in: Krantz, M. Keeping tabs online: doing business on the Net is hard because the underlying software is so dumb. XML will fix that. *Time Online*, 10 November 1997. http://www.pathfinder.com/time/magazine/1997/dom/971110/technology.keeping_tabs_.html

30. Bosak, J. and Bray, T. XML and the second-generation web. *Scientific American*, May 1999. http://www.sciam.com/1999/0599issue/0599bosak.html

31. Bosak, J. Media-independant publishing: four myths about XML. *IEEE Computer*, 31(10), 1998, 120–120.

32. Johnson, M. XML for the absolute beginner: a guided tour from HTML to processing XML with Java. *Java World*, April 1999. http://www.javaworld.com/javaworld/jw-04–1999/jw-04–xml_p.html

33. Matsumura, M. XML speeds along in standards land. *Java World*, February 1998. http: //wwwjavaworld.com/javaworld/jw-02–1998/jw-02–miko.html

34. Connolly, D., Khare, R. and Rifkin, A. The evolution of web documents. XML.com. *XML.com*, 2 October 1997. http://www.xml.com/xml/pub/w3j/s3.connolly.html

35. Bray, T. The trouble with browsers. *XML.com*, 18 December 1998. http://www.xml.com/xml/pub/98/12/B-Trouble.html

36. Reichard, K. XML: a cross-platform tool for e-commerce. *Unixintegration.com*, September 1998. http://www.performancecomputing.com/unixintegration/9809/9809f1.htm

37. Rath, H.H. XML: chance and challenge for online information providers. *The SGML/XML Web Page*, [no date]. http://www.oasis-open.org/html/xml__chance_challenge.html

38. Freter, T. XML: document and information management. *Sun Microsystems*, September 1998. http://www.sun.com/980908/xml/

39. Rosenfeld, J. Quoted in: Oberndorf, S. XML: a new Internet language. *Catalog Age Magazine*, 1 April 1999. http://www.catalogagemag.com/Magazines/CatalogAge/9904/1999040106.html

40. Glave, J. One man's vision of the Web's future. *Hotwired*, 15 April 1998. http://www.wired.com/news/news/technology/story/11693.html

41. Shimmin, B.F. XML reality must include openness. *LAN Times*, June 1998.http://www.lantimes.com/98/98jun/806a025b.html

42. Gibbs, W.W. The Web learns to read. *Scientific American*, June 1998. http://www.sciam.com/1998/0698issue/0698cyber.html

43. Trommer, D. XML lets buyers comparison-shop. *EBN News*, October 1998. http://www.ebnews.com/section/102698/online2.html

44. Mougayar, W. Web agent will give users some leverage. *LAN Times*, November 1997. http://www.lantimes.com/97/97nov/711b041b.html

45. Meserve, J.J. Searching for the Holy Grail, XML style. *Application Development Trends*, August 1998. http://www.adtmag.com/Pub/Aug98/fe804.htm

46. Bremser, W. First baby steps on the journey to working XML. *Internet World*, 30 March 1998. http://www.internetworld.com/print/1998/03/30/iwlabs/19980330–baby.html

47. Walsh, J. XML explodes onto the scene: vendors rush to release tools. *Network Computing*, 30 November 1998. http://www.networkcomputing.com/1013/1013colwalsh.html

48. Bosak, J. XML, Java, and the future of the Web. *Sun Microsystems*, 3 October 1997. http://metalab.unc.edu/pub/sun-info/standards/xml/why/xmlapps.htm

49. Schaeffer, B. Tracking the emerging impact of XML. *X-Systems*, 1 November 1998. http://www.strategies.com/xml1198.htm

50. Denison, D.C. The road to XML: adapting SGML to the Web. *XML.com*, 2 October 1997. http://www.xml.com/xml/pub/w3j/s1.discussion.html

51. St. Laurent, S. *Cement shoes for XML?* Author's web site, 1998. http://www.simonstl.com/articles/cement.htm

52. Bray, T. Quoted in: Maclachlan, M. XML will make life simpler, editor says. *Techweb*, 12 March 1999. http://www.techweb.com/wire/story/TWB19990312S0024

Embarking on the Internet: what motivates people?

Reijo Savolainen
Department of Information Studies, FIN-33014 University of Tampere, Finland
liresa@uta.fi

Issues of becoming an Internet user primarily in non-work contexts are discussed. The author reviews individual, social and economic factors motivating network use. Main findings of an empirical study based on the interviews of 23 Finns in 1997 are analysed. In addition, comparisons are made with the findings of the Consumer Research Project (1996–98), conducted in Finland. The main motivating factors leading to network use were the requirements of work and study; also making everyday transactions easier, for example by home banking, was seen as important. The network connections were also acquired for keeping in contact with others and for hobbies, whereas entertainment purposes were not valued.

Introduction

The surveys reporting the rapidly growing use of the Internet rarely discuss the motivations behind becoming a network user. However, the motivational issues are central if we explore, for example, the reasons for the 'digital divide' recently noticed, e.g. in the United States [1]. In the early stages of Internet diffusion, the major group of users consisted of young males with a relatively high level of education [2]. In the following phases of diffusion, the differences of sociodemographic features have tended to be levelled off, even though quite slowly.

As to the growth of Internet use, Finland offers an interesting case. Since the middle of the 1990s, Finns can be counted among the precursors taking the Internet into daily use [3]. Thus, the experiences obtained in Finland may elicit interest when the future penetration of networked services to households is considered in an international perspective.

Based on empirical studies conducted in Finland, the present paper asks what makes people interested in the Internet, particularly in non-work contexts where people consider the acquisition of network connections at home. Individual, social and economic factors triggering Internet use will be discussed. The major attention will be devoted to people's expectations towards the networked services and the potential needs behind the utilisation of the Internet. Taking this viewpoint, the needs of the established users will not be discussed in the present paper [4]. Comparative notions will be made to the findings of the Consumer Research Project conducted in 1996–98. The project explored, among others, the adoption of email, WWW and Internet-TV in Finnish households.

The attraction of the Internet

Even though the Internet is becoming an everyday media, it still has a novel charm. The Internet can be seen as a new medium that holds exciting promises in regard to information seeking, communication, entertainment and home shopping. The prospects of utilising the Internet via WAP mobile telephone may strengthen this image. In terms of Ithiel de Sola Pool [5], the electronic networks exemplify 'technologies of freedom'; they broaden the repertoire of communication channels, thus decreasing one's dependence on traditional sources of information, for example, the main newspapers. On the other hand, the media tend to give a positive picture of the networks: the Internet epitomises the advent of the new millennium, offering the major forum of the future information society. Obviously, 'the spirit of our time' favours one's considerations to become a network user.

Notwithstanding the hype associated with new media, the attractiveness of the Internet seems to be largely based on its multifunctionality: the same tool can be used for several purposes. According to Mark Stefik [6], the attractiveness of the Internet can be explained by the user's ability to adopt archetypal roles such as communicator, trader and adventurer. For the first time in the history of media, an individual may take the above roles at the same forum: he or she may communicate by email across the continents, buy and sell products at the global e-markets and search for thrilling experiences in the jungle of WWW pages.

There are a number of other cultural and social factors motivating people to start using the Internet. First, the predominating values of society may affect, even though indirectly, the ways in which people assess the significance of computers and the Internet. In recent years, the importance of 'information superhighways' has been emphasised in the strategic plans for the development of the information society, published by national governments and the EU. For example, in strategic papers published in Finland, one may find 'rhetorics of responsibility' suggesting that citizens should become active users of the Internet in order to be competitive in the information society [7]. Network skills become more important because increasingly, the work tasks will become information intensive. Castells [8] crystallises this trend by asserting that the future society will be characterised by 'informational capitalism' on a global level and that the Internet forms the major infrastructure of this kind of society. In this perspective, the job-related utilisation of the Internet will no longer be a question of free choice but rather that regular use becomes a necessity for competitive organisations. However, the home use of the Internet may be less compelling because it tends to be motivated by hobbies and other non-work activities.

To accept or to reject?

Personal expectations, attitudes and preferences are central when the motivating factors of becoming an Internet user at home are discussed. The acceptance and utilisation of novel products and services can be explained by several models primarily developed in psychology, theories of innovation diffusion and communication research [9]. The psychology of user acceptance can be made understandable, for example, by the Theory Reasoned Action (TRA) [10]. It suggests that a person's behaviour is determined by his or her intention to perform the behaviour. His or her attitudes and beliefs concerning the consequences of action influence the intention. In addition, the intention is affected by normative beliefs of other people, for example, friends or colleagues.

Based on the ideas of TRA, Davis *et al* [11] proposes a Technology Acceptance Model (TAM). For example, the utilisation of the computer is explained by the behavioural intentions to use and attitudes toward using. These factors are determined by perceived usefulness and perceived ease of use; in addition, there may be external variables influencing the above factors. Even though TRA and TAM offer an interesting explanation of the factors determining use, the models remain quite abstract. They shed light on the general factors triggering the use but fail to discuss the substantive reasons why people who have not utilised the Internet before may become network users. The theories of innovation diffusion offer more concrete answers to the above statement. Rogers [12] identified five factors that explain why people accept or reject innovations such as the Internet:

1. Relative advantage: to what extent the innovation is conceived an improvement on products or services available thus far? The bigger the advantage, the more rapidly the innovation will be adopted.
2. Compatibility with values, beliefs, practices, experiences and need: the better the compatibility, the easier to accept the innovation.
3. Complexity: the more an innovation requires new skills, the more slowly it will be diffused to the user society.
4. Trialability: the diffusion of an innovation will be facilitated if the novel product or service can be tried without risks and if positive experiences are received from use.

5. Observability of gains: the more evident the gains received from use, the easier to accept the innovation.

According to Rogers [12], the people first adopting an innovation form the group of 'innovators'. Often for curiosity, they take the novel products or services into use when they appear at the marketplace. Rogers suggested that only about 2 per cent of the population might be classified as this avant-garde. The ensuing group, 'early adopters', is somewhat larger, about 10 per cent of the population. For example, in Finland, the regular use of the Internet has already expanded to this group. In the third phase, the 'early majority' enters the field. At the end of this phase, half of the population can be counted as users of the innovation. The fourth group consists of the 'late majority' (about a third of the population); finally, the 'laggards' follow suit. Regarding the penetration rates of radio and TV in Western countries, the 'laggards' have joined the party. In the case of the Internet, the corresponding development might take some 10–20 years, depending, for example, on the popularity of the Internet-TV.

Becoming a network user at home

The above discussion suggests that the decision to start network use at home is affected by a number of social and individual factors. The fashionableness of the Internet may not suffice alone but positive attitudes and expectations towards the Internet, as well as an interest in the networks, are required for a favourable decision. The expectations are affected by one's belief of how the use of the Internet may, for example, facilitate keeping in contact with friends living abroad. Further, one may consider, how enjoyable the use will be and if there will be competing needs for the time spent surfing on the Net [13]. Thus, the use of the Internet should be in concert with one's 'way of life', in particular the prefences of free-time activities [14].

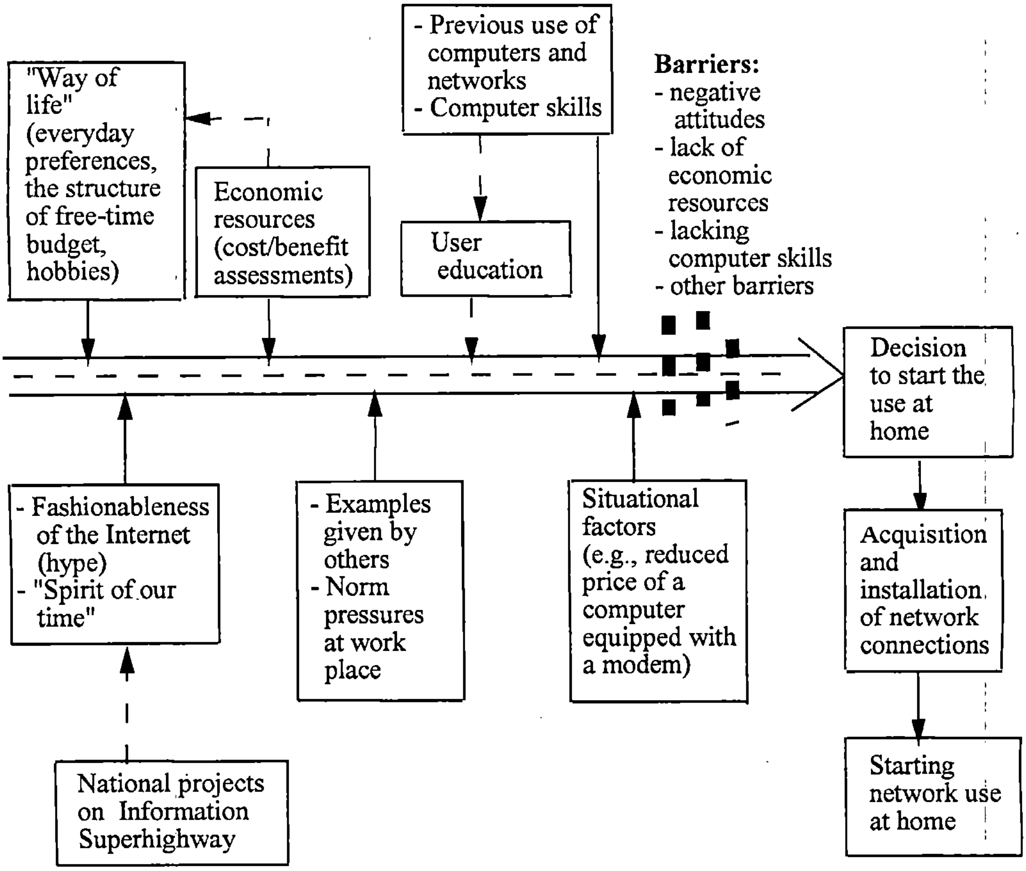
In addition, the costs involved in the acquisition and ongoing use of the Internet should not exceed the benefits to be expected. Hence, affordable access is one of the basic criteria when network use at home is considered. At a household level we face here the issues of a 'moral economy': how important are the network connections for the family and what are the benefits which make this investment reasonable [15]. The Internet should offer at least one relative advantage, compared to that of earlier ways of doing things, preferably in a long-term perspective. If the relative advantages are not evident enough, the pleasure of using a novel service *per se* may not suffice as a motivation for use [16]. Further, the valuations of friends and colleagues may affect the considerations of starting the use: for example, whether friends welcome email messages compared to the telephone [17]. Naturally, becoming a network user requires the acquisition of necessary network skills. This requirement may be found particularly demanding among elderly people not familiar with computers themselves.

If the basic requirements are not met sufficiently, becoming a home user is improbable. Negative attitudes towards computers and deeply ingrained media habits may form the highest threshold. Thus, one may believe that, compared to other media, the networked services have nothing new to offer. The Internet may be conceived, for example, as a fashionable channel primarily distributing pornography. Parents may worry about the potential harm caused by the Internet, thus refusing to acquire network connections in order to protect their children from obscene material. In some cases, the lack of computer skills may form a significant barrier. The threshold may be conceived even higher if it is believed that the learning of network skills requires a long-time wrestling with computer manuals. Moreover, economic barriers may be felt too high. For example, unemployed people may not be able to invest £1000 for a computer and modem plus paying considerable sums for monthly use. On the other hand, if the user is really motivated, the required investments will be made even though they may take a considerable share of a person's or a household's monthly budget.

The most central factors affecting the decision to become a network user are illuminated in Figure 1.

The importance of the factors presented

Figure 1. *Becoming a network user.*


"Way of life" (everyday preferences, the structure of free-time budget, hobbies)
Economic resources (cost/benefit assessments)
- Previous use of computers and networks
- Computer skills
User education
Barriers:
- negative attitudes
- lack of economic resources
- lacking computer skills
- other barriers
Decision to start the use at home
Acquisition and installation of network connections
Starting network use at home
- Fashionableness of the Internet (hype)
- "Spirit of our time"
- Examples given by others
- Norm pressures at work place
Situational factors (e.g., reduced price of a computer equipped with a modem)
National projects on Information Superhighway

in Figure 1 varies from one individual to another. Thus, it is not possible to identify the most central factors explaining why an individual becomes a network user. The arrow moving from left to right stands for the considerations that precede the decision to start Internet use at home. Social and cultural factors exert indirect influence on these considerations; these factors manifest themselves, for example, in the fashionableness of the Internet. In addition, the considerations regarding use may be affected by examples given by friends and colleagues. On the individual level, the 'way of life', as a constellation of everyday preferences and habits, is central because the 'way of life' reflects one's major interests manifesting themselves, for example, in hobbies and the structure of free-time budget. In the economic considerations, the affordability of Internet use plays an equally central role. If the considerations associated with the above factors do not result in a clear preference for or against use, situational factors may be decisive. An opportunity to acquire network connections at a reduced price may exemplify this kind of situation, leading one to decide in favour of home use. However, barriers such as poor computer skills or low income may affect the considerations from the beginning, and the plans to start home use may be abandoned for the present.

Empirical data

The empirical data of the present study was gathered by theme interviews in spring 1997, as a part of a larger study focusing on the ways in which Finns use the Internet. The major themes focused on the ways in which citizens utilise the networked services such as email, WWW and discussion groups; in addition, the problems faced in use were discussed [14, 18].

Altogether 23 people living in the region of Tampere, Finland, participated in the interviews conducted by the author. Of the interviewees, 8 were women and 15 men. On average, an interviewee was 32 years old. Non-users or potential users were not recruited to the study. At the time of the interviews, 11 of the interviewees had acquired network connections at home. Most of the interviewees had the opportunity to use the Internet at their workplace or at school. Only two were dependent on the networked services offered in public libraries. A couple had already started network use in the late 1980s when the commercial videotext services became available in Finland. At that time, the services were primarily used for home banking. Because all the interviewees were Internet users, the assessments concerning the expectations and motivations towards becoming a network user were based on retrospective views. However, because most interviewees had become users not earlier than two to three years prior to the interview, they were able to recall the major factors that elicited their interest towards the Internet.

In addition, comparative notions are made to the findings of the national Consumer Research Project (1996–98) which surveyed Finnish people's attitudes and expectations towards different network services [16, 19]. The data from the above-mentioned project is based on the interviews of about 300 persons. In addition, the data include a nation-wide telephone survey conducted in autumn 1998, about 2000 persons aged 15–74 participated in the study [20].

Expectations and motivations of use

Earlier studies conducted in Sweden [21] and in the United States [22] have revealed that the major reasons leading to network use at home tend to be pragmatic. In a representative survey of 2500 Americans Katz and Aspden [22] found that the major attractions of Internet use were the opportunity to communicate by email and find information of special interest. Broadly speaking, the studies conducted in Finland resulted in similar findings even though there were some differences.

The interviews conducted by the author revealed several reasons for starting network use. Table 1 summarises the major motivating factors concerning both job-related and non-work use. Because most interviewees reported multiple reasons, the number of motivations exceeded the number of participants (23). Due to the low number of informants, the results cannot be generalised statistically. However, the findings are indicative, as far as the popularity of major motivations to become an Internet user is concerned.

The interviews revealed that people rarely have acute needs, which lead to acquiring network connections at home. In most cases, the network connections were acquired for serious reasons, not for enter-

Table 1. *The major reasons for starting Internet use.*

Reasons for starting Internet use	*n*	%
the requirements of work and study	9	17.3
saving time, easing of daily transactions	7	13.5
curiosity	6	11.5
facilitating contacts	6	11.5
hobbies	5	9.6
Internet courses at workplace	5	9.6
examples given by others	3	5.8
computers as a hobby	3	5.8
enhancing information seeking	3	5.8
the fashionableness of the Internet	2	3.8
facilitating international contacts	2	3.8
economic benefits	1	1.9
Total	**52**	**99.9**

tainment. The requirements of work and study appeared to be the most popular reasons and they originated from job-related contexts. For example, studying at university necessitated the utilisation of email and WWW. Also the Internet courses provided at the work place served as a source of inspiration.

Most interviewees had familiarised themselves with computers at the workplace or at school. In some cases, short courses on the use of the Internet were offered. Even though the examples given by friends and colleagues might have affected the decision to start use, most interviewees pointed out that their own considerations were most central in this respect. Only a couple of the informants mentioned the indirect influence of other people:

> A network 'guru' was invited to talk about the Internet at our work place and before the presentation I was not particularly interested in this issue. However, I went there and suddenly I realised that all kind of things could be found from the Net. The presentation really aroused my interest and besides, I had some colleagues who already were users and they told me more about the Internet.

The above motivation is closely related to the need to keep abreast of the developments occurring in the field of computers and electronic networks. To be able to meet the changing requirements of working life and in general, the challenges of the information society, the acquisition of network connections may be seen as indispensable. This motivation may also be linked with the need to provide opportunities for children to learn computer and network skills at home because that would give them a head start in life. The positive experiences received from use at the work place elicited interest to acquire network connections at home, too.

> I have used the Internet for a longer time at the work place and found that the Net is a treasure trove in many respects. For example, you may retrieve facts, official papers, memos, electronic versions of newspapers and so on. When my colleague told about the printouts he had taken from the Web pages I realised that actually, the Internet is an archive or a library. Of course there is also all kinds of garbage but I am not interested in it. I started with searching for the material I need for my job. One day I noticed that even real time sports news from the Olympic Games is available on the Web. So I thought it would be nice to have the Internet at home. Then I noticed an advertisement of a cheap ISDN modem in a newspaper and that decided all.

Thinking of the non-work related motivations, the opportunity to ease everyday transactions, for example, by means of home banking was popular. The relative advantages particularly brought by time-saving were seen as considerable, because the need for frequent visits to bank offices could be eliminated. Also the Consumer Research Project [16, 20] revealed that the opportunity to ease everyday transactions by means of home banking is one of the major reasons. On the other hand, electronic shopping did not elicit similar interest [16]. However, the attractiveness of home shopping may increase in the future when the provision of goods and services will proliferate on the Net.

Contrary to the findings obtained in the United States, the Finnish would-be users participating in the consumer study did not expect that the Internet would offer new prospects of making easy money [16]. This was also evident in the interviews conducted by the author. In the survey conducted by Katz and Aspden [22] in the United States, the motive of 'business opportunity to make money' was highly popular among those having heard of the Internet but not yet using it. However, the same study indicated that the actual users had realised quite soon the limited opportunities of doing business in the new marketplace. The Net is not necessarily a 'gold mine', even though in favourable cases the advertisements distributed through the Web may reach millions of potential buyers all over the world.

As Table 1 indicates, the opportunity to facilitate everyday contacts by email was valued highly among the interviewees participating in the study. Email was perceived as an expedient and cheap way to keep in contact with friends:

> I have friends studying all over the country and sometimes abroad. Now it is much easier to contact them; I am not very fond of the telephone but email is very nice.

Novice users often believe that the Internet will provide a channel for real-time interactive communication. However, those having used the Internet for a long time were well aware of the fact that the messages will not necessarily reach the recipient in a few seconds, because he or she is not available or the network connections do not work. For most interviewees, email was the first networked service to be utilised; email also elicited interest towards other services, particularly WWW. Email was also mentioned when the opportunities to replace other media, for example, telephone and snail mail was discussed. However, the possibility of substituting services for others was not emphasised.

Some interviewees mentioned the prospect of enhancing information seeking. The following excerpt from an interview illustrates the expectations towards the Internet as an information source:

> I thought that there would also be economic benefits because I live quite far away from the town and it takes time to visit the library. Having the network connections at home would lower the threshold to seek information. For example, at the moment I have to use medicine and I am interested in the potential side effects of the drug. Naturally, I may call the drugstore but it might be easier to check the Internet.

However, when the first searches from the Internet were made, many of interviewees realised that the expectations were not necessarily met. Some of the interviewees participating in the Consumer Research Project had expected that the WWW pages would provide a kind of 'data bank' for use (16). However, they learned quite soon that the quality of information available on the Web varies considerably. Similarly, the electronic versions of newspapers were assessed more favourably before the use. Those having familiarised themselves with the electronic versions felt that they provided less information than the printed ones and that the electronic versions tended to favour entertainment.

These findings indicated – as those of the Consumer Research Project [19] did – that the need for the networked services will be developed on the basis of one's vague expectations because there is not (as yet) enough real use experiences. Thus, motivation is affected by the user's perceptions concerning the qualities of the media: what kind of relative advantages the Internet might be able to provide? Naturally, the perceptions of alternative media also affect one's interpretation and reinterpretation of needs.

As to the other reasons for becoming a network user, curiosity was mentioned relatively often among the informants, even though it was rarely seen as the major motive. Similarly, the fashionableness of the Internet attracted some of the youngest informants. Other reasons listed in Table 1 are dispersed to a rather broad area and they were associated with non-work issues such as hobbies, for example, writing fiction.

Is there a need for Internet-TV?

The National Consumer Research Project indicated that if the computer is not seen as useful for the family, the purchase of Internet-TV was not perceived as meaningful (19). Perhaps this is due to the fact that the possibilities to utilise computers seem to be broader than those of Internet-TV. Internet-TV seems to serve best those who cannot afford a computer with a modem, or those having no need for network connections at home because the Internet is available at the workplace, free of charge. However, the acquisition of Internet-TV may be supported by the fact that most people have already familiarised themselves with TV and thus, the resistance of hybrid products such as the Internet-TV may be weaker.

The expected use of Internet-TV was primarily associated with entertainment: the services will be utilised in the living room, lying on the sofa. In general, the need for Internet-TV was not seen as particularly acute or burning. This view reflects the more general valuations adopted by the Finnish network users. As noted above, Finns prefer the serious uses of the Internet for the purposes of work and study. The nation-wide study of the Consumer Research Project revealed that, in 1998, only 2 per cent of Internet users felt that entertainment, such as games and erotic services, available on the Web are really important [23].

In the above survey, other reasons why people are not interested in acquiring Internet connections at home were explored [20]. Some interviewees asserted that their home computer was too old for the installation of Internet connections. However, the most frequent reason was simply that there is no need for the Internet at home. This might be due to a total lack of interest, or the opportunity to utilise the networks elsewhere, for example, at school or work-place. Even though one could afford the acquisition of network connections, they will not be purchased merely for status reasons. Thus, positive attitudes based on favourable cost/benefit considerations form the necessary condition when the acquisition of connections are weighed up, particularly in cases where the income level of the household is low and there is lot of competing needs. However, further research is needed on the relative importance of the economic and attitudinal factors affecting the decision to accept or reject the Internet.

Conclusions

The present paper focused on the views of the 'early adopters' of the Internet. The main motivating factors leading to network use appeared to be the requirements of work and study. Also communicational factors were popular: email was conceived as a new way to keep in contact with other people. The network connections were also purchased for home banking, hobbies and information seeking. In contrast, entertainment purposes alone do not seem to motivate people in this respect.

Becoming a network user can be explained by a number of individual, social and economic factors that are intertwined. In addition, the would-be user has to acquire minimum skills of use or have the motivation to learn them. Social factors manifest themselves in the examples received from others. Economic reasons refer to the individual's (or family's) ability to acquire network connections and pay the costs of ongoing use. At the level of the family, these questions often boil down to the issues of 'moral economy': how important it is to purchase network connections and what are the expected benefits which makes this investment reasonable? Further, the use of the Internet should be in accord with one's 'way of life'. The hours spent on network use means less time for other purposes, for example, watching TV or outdoor recreation. Hence, the Internet has to occupy its place in people's time budgets. In addition, the use should bring relative advantages, which manifest themselves, for example, in time saving and the easing of everyday transactions.

The studies focusing on Internet use thus far have rarely discussed the motivations of becoming a user. The findings of the present paper are based on retrospective views of people already using the Internet. In order to finish the picture, the attitudes and expectations of would-be users, or not-yet-users, should be explored as a real-time process. Studies such as these are relevant because, in most countries, the majority of people still belong to the group of non-users. The topic is also worth studying because the motivations of the 'late majority' and 'laggards' may differ from those of 'early adopters' and 'early majority', which currently dominate the field of Internet users.

References

1. National Telecommunications and Information Administration (NTIA). *Falling through the net: defining the digital divide.* http://www.ntia.doc.gov/ntiahome/fttn99 (visited 4 February 2000).

2. Atkin, D., Jeffres, L.W. and Neuendorf, K.A. Understanding Internet adoption as telecommunication behavior. *Journal of Broadcasting and Electronic Media*, 42(4), 1998, 475–490.

3. NUA. *How many online? Europe.* http://www.

nua.ie/surveys/how_many_online/ europe.html (visited 4 February 2000).

4. Kraut, R. *et al. Why people use the Internet.* http://homenet.hcii.cs.cmu.edu/ progress/purpose.html (visited 4 February 2000).

5. de Sola Pool, I. *Technologies of freedom.* Cambridge, Mass.: Harvard University Press, 1983.

6. Stefik, M. *Internet dreams. Archetypes, myths and metaphors.* Cambridge, Mass.: MIT Press, 1996.

7. Sitra. *Elämänlaatu, osaaminen ja kilpailukyky. Tietoyhteiskunnan strategisen kehittämisen lähtökohdat ja päämäärät* [Quality of life, knowledge and competitiveness. The starting points and goals of the strategic development of the information society]. Helsinki: Sitra, 1998. (Sitra Publications 206)

8. Castells, M. *The information age. Economy, society and culture. Volume 1. The rise of the networked society.* Oxford: Blackwells, 1996.

9. Dillon, A. and Morris, M.G. User acceptance of information technology: theories and models. In: Williams, M.E., ed. *Annual Review of Information Science and Technology.* Volume 31. Medford, NJ: Information Today, Inc., 1996, 3–32.

10. Fishbein, M. and Ajzen, I. *Belief, attitude, intention and behavior. An introduction to theory and research.* Reading, MA: Addison-Wesley, 1975.

11. Davis, F. *et al.* User acceptance of computer technology. A comparison of two theoretical models. *Management Science*, 35(8), 1989, 982–1003.

12. Rogers, E. *Diffusion of innovations.* 3rd edition. New York: Free Press, 1983.

13. Klobas, J.E. Beyond information quality: fitness for purpose and electronic information resource use. *Journal of Information Science*, 21(2), 1995, 95–114.

14. Savolainen, R. Seeking and using information from the Internet. The context of non-work use. In: Wilson, T.D. and Allen, D.K., eds. *Exploring the contexts of information behaviour.* London: Taylor Graham, 1999, 356–370.

15. Silverstone, R. *et al.* Information and communication technologies and the moral economy of the household. In: Silverstone, R. and Hirsch, E., eds. *Consuming technologies. Media and information in domestic spaces.* London: Routledge 1992, 15–31.

16. Puirava, M., ed. *Kuluttajat ja multimediapalvelut* [Consumers and multimedia services]. Helsinki: National Technology Agency (Tekes), 1997. (Reports of Digital Media, 1 January 1997.)

17. Fulk, J. *et al.* The dynamics of context-behaviour interactions in computer-mediated communication. In: Lea, M., ed. *Contexts of computer-mediated communication.* New York: Harvester Wheatsheaf, 1992, 7–29.

18. Savolainen, R. The role of the Internet in information seeking. Putting the networked services in context. *Information Processing & Management*, 35(6), 1999, 765–782.

19. Turpeinen, P. Tarpeet ja motiivit [Needs and motives]. In: *Uusmedia kuluttajan silmin* [New media through consumer's eyes]. Helsinki: National Technology Agency (Tekes), 1998, 51–86. (Reports of Digital Media 2/1998)

20. Marttila, M. Tietokoneen ja Internetin hankkiminen kotiin [Acquiring computer and Internet connection for home use]. In: *Internet suomalaisten arjessa* [The Internet in the everyday life of Finns]. Helsinki: National Technology Agency (Tekes), 1999, 43–60. (Reports of Digital Media 3/1999)

21. Christiansen, B. *IT-brug i hverdagslivet. En kvalitativ interviewundersøgelse om ti menneskers computer- og Internetbrug i privatsfären* [The use of information technology in the context of everyday life. A qualitative interview study of ten persons using computers and the Internet in the private sphere]. Gothenburg: University of Gothenburg, 1996. (Research Report 66)

22. Katz, J. and Aspden, P. Motives, hurdles and dropouts. *Communications of the ACM*, 10(4), 1997, 97–102.

23. Turpeinen, P. Internetin käyttömotiivit [The motives to use the Internet]. In: *Internet suomalaisten arjessa* [The Internet in the everyday life of Finns]. Helsinki: National Technology Agency (Tekes), 1999, 61–80. (Reports of Digital Media 3/1999)

Recent titles from Aslib

Guide to business information on Central and Eastern Europe

Tania Konn

234 × 156mm March 2000 vi, 235pp
0 85142 428 7 paperback £35.00 (£28.00 Aslib corporate members)

This guide is an introduction to English language sources, in electronic and conventional print forms, dealing with Central and Eastern European business issues. It gives evaluative descriptions and costs of all listed sources, and concentrates on recent sources. Sources in respect of some of these countries can be difficult to locate, and the author provides guidance on how to go about finding them.

Contents: Under each country, information sources are grouped in broad categories: Overview (sources designed to answer general, exploratory, country and regional questions, e.g. population, politics, how to do business etc); Current developments (e.g. recent changes in tax and other laws, trends in foreign direct investment, latest project tenders); Companies and contacts; Industries and services; Legislation; Organisations (a listing of agencies and bodies able to provide assistance, information and data to business people). Index.

Covers: Albania; Bosnia-Herzegovina; Bulgaria; Croatia; Czech Republic; Hungary; Macedonia; Poland; Romania; Serbia and Montenegro; Slovenia.

i in the sky: visions of the information future

Edited by Alison Scammell

234 × 156mm December 1999 xii, 307pp
0 85142 431 7 Paperback £37.50 (£30 Aslib corporate members)

i in the sky is a collection of essays by over 40 experts, including leading writers Charles Handy and Don Tapscott, giving their personal vision of the future of information. Information here is given its widest meaning and includes such subjects as electronic commerce, cybernetics, robotics, artificial intelligence, and even computers as fashion accessories. Information as a phenomenon pervades all areas of life and its evolution has consequences for everyone. Many of the papers have central themes, such as the future of computer intelligence, library and information services, interactive Internet marketing, and networked learning in higher education.

Aslib imi
INFORMATION MANAGEMENT

Information insights: case studies in information management

Edited by Sylvia Simmons

234 × 156mm December 1999 viii, 176pp
0 85142 427 9 paperback £37.50 (£30 Aslib corporate members)

This book features updated and revised versions of interviews previously published in Aslib's monthly magazine, *Managing Information*, providing revealing, lively and forward-looking snapshots of a wide range of information services – from the scholarly collection to the global focus of a worldwide knowledge resource centre. The Editor has brought together the various themes arising from the interviews and emphasised the practical implications for all information managers. These case studies from leading organisations show clearly what information managers can achieve, and how many have managed to re-invent themselves and place information and information professionals where they belong – at the very heart of their company's business and future success.

Case studies include: Amnesty International; the Boots Company; BP Amoco; British Airways; the BBC; the British Film Institute; the Chartered Institute of Marketing; Deloitte & Touche; the Department of Social Security; the Economist Intelligence Unit; the European Bank for Reconstruction and Development; *The Guardian / The Observer*; Guinness Archives and Records Centre; the Metropolitan Police Service Library; Motorola; the National Museums of Scotland Library; St Georges Hospital Medical School; SmithKline Beecham; Sutton Library Service; the Tate Gallery; UNESCO; Xerox Europe Technical Centre.

The Internet: its impact and evaluation

Edited by David Nicholas and Ian Rowlands

234 x 156mm June 2000 200pp approx
0 85142 438 4 paperback £39.99 (£32 Aslib corporate members)

The Proceedings of an international forum held in 1999, whose purpose was to: explore methods for evaluating the Internet's many features and functions; assess the impact of the Internet in strategic areas of human endeavour, so providing us with an understanding of the way that the Internet is fashioning the world we live in; identify the key information and communication principles, ideas and concepts that are emerging as a result of the spread and development of the Internet; and put down an agenda for future Internet research and collaboration.

Contents: The challenge: unstable knowledge, unstable times; Three-dimensional impact: a case study of electronic government; Joining up information and learning; Newspapers and the Net: peaceful coexistence or fight to the death? The impact of the Internet on the ownership of information; Impact of the Internet: some conceptual and methodological issues; Studying the impacts of the Internet without assuming technological determinism; The relevance of information retrieval research for the evaluation of the WWW; From bibliometrics to cybermetrics; Web metrics: too much data, too little analysis; Surveying the global diffusion of the Internet; Tracking and evaluating the global information consumer; Evaluating the net as a 'local' information resource.

Aslib imi
INFORMATION MANAGEMENT

Strategic management of information services: a planning handbook

Sheila Corrall

234 x 156mm June 2000 380pp approx
0 85142 346 9 paperback £39.99 (£31.99 Aslib corporate members)

This book provides an in-depth analysis of strategic management concepts and techniques and how they can be usefully applied to the planning and delivery of information services. It offers practical guidance on the whole strategy process, from appraisal and assessment through to implementation and improvement. Annotated references to management and information service literature enable the reader to pursue chosen topics in more depth.

Partial contents: Planning paradigms; Environmental issues; Strategic focus; Strategy formation; Money matters; Achieving change; Securing capability; Managing flexibility; Appendices; Index.

ORDER FORM

Please photocopy the form below, complete it, and return to:

Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, UK
Tel: +44 (0) 1206 796 351 Fax: +44 (0) 1206 799 331 Email: sales@portlandpress.com

Please send me:

Title	ISBN	Price	Qty	£
Guide to business information on Central and Eastern Europe	0 85142 428 7	£35 £28*	☐ ☐	☐ ☐
i in the sky: visions of the information future	0 85142 431 7	£37.50 £30*	☐ ☐	☐ ☐
Information insights: case studies in information management	0 85142 427 9	£37.50 £30*	☐ ☐	☐ ☐
The Internet: its impact and evaluation	0 85142 438 4	£39.99 £32*	☐ ☐	☐ ☐
Strategic management of information services: a planning handbook	0 85142 346 9	£39.99 £31.99*	☐ ☐	☐ ☐

*Aslib corporate members' price

Please tick as appropriate

☐ Please send me a pro forma invoice

☐ I enclose a cheque/bank draft payable to: Portland Press Ltd for £*

☐ Debit my Mastercard/Visa/Amex for £ ____________

Signature ____________

Expiry date ____________

Please add £6 ($10) to non-sterling payments and cheques drawn on non-UK banks

Name ____________

Position ____________

Organisation ____________

Address ____________

Postcode ____________

Tel ____________ Fax ____________

Email ____________

**Postage: £2.50 per book up to a maximum of £7.50 for 3 or more in any single order*

INFORMATION MANAGEMENT

CONTENTS

197–199 Comment
Mobile phones: the new information medium?
Tom Dobrowolski, David Nicholas and Jonathan Raper

200–206 British official publishing in the Internet Age
Chris Pond

207–217 An observational study of the information seeking behaviour of Members of Parliament in the United Kingdom
Robert Orton, Rita Marcella and Graeme Baxter

218–231 A user-centred approach to the evaluation of Subject Based Information Gateways: case study SOSIG
Maria Monopoli and David Nicholas

Comment
Mobile phones: the new information medium?

Tom Dobrowolski[1]*, David Nicholas and Jonathan Raper*
Infomobility Study Group[2]*, Department of Information Science, City University, Northampton Square, London EC1V 0HB*
Whtom@eudoramail.com

Introduction

We have all been lulled into thinking that the personal computer is the natural platform for information seeking activities of all kinds. The arrival of the Web surely cemented its position, because, via the Web, it is possible to obtain vast amounts of information in text, pictorial and audio forms, and the opportunities for processing and packaging this data are endless. Of course, a price has to be paid in terms of user-friendliness/simplicity of use – computers are still not easy to use, only a small percentage of their functions are ever utilised and their mass market credentials are questionable. Not so long ago it did seem on the cards that television – in the form of web access or digital TV – would pose something of a threat to the PC. But the threat appears to be a long time coming and with the passage of time the threat appears to be dissipating.

So, what do we make of the digital mobile phone's sudden appearance on the information seeking and retrieval stage[3]? Agreed, that in some ways the mobile phone always has been part of the scene, but only really as a bit player, whose domain is really oral and informal information. But now we are talking WAP, Amazon.com, and location-sensitive information services. Moreover we are talking about the likes of Nokia, Vodafone and not Microsoft and Dialog. Phones connected to computers and the Web. Phones dispensing cans of Pepsi. According to Vodafone phones are already engaged in the transfer of some 40 million textual messages a month. Something big has happened. This is now serious and heady stuff for information professionals.

Mobile phones challenge current IS thinking

If the phone is to become *the* platform for information seeking then this will represent a fundamental re-think of information design, provision and use. Phones will not deliver/display large volumes of data, or handle large amounts of pictures and graphics – text will be what it dispenses, and it will be in small doses (largely financial, shopping, sports and local information to begin with). However, the new breed of colour screen personal organisers coming this year with phone and satellite navigation built-in will be used to download the Daily Me, to plot maps of current location and to read the next generation of e-books. These services will come via the next information killer-apps- micro-browsers such as Palmscape. Prioritising information will be the key skill in the mobile information age. That is one that journalists' possess but which information professionals will have to master.

---

[1] University of Warsaw member.

[2] Currently this group of information professionals and geographical information scientists are investigating the delivery and use of news and health information over the mobile phone.

[3] At February's world computer fair CEBIT 2000 (Hannover, Germany) all of the world's mobile phone makers demonstrated phones capable of Internet access.

If computers are all about interfaces, complex retrieval and massive storage, mobile phones are all about communication, transparency, simplicity and, most importantly, human contact. In fact, does not the phone represent the perfect information-seeking platform, combining as it does the best of the human and computer information worlds? The interplay between human and machine information delivery is surely a powerful one; the choice between a human and machine interface is similarly attractive. There is also the coming together of the informal and formal information worlds, the text or imagery and oral communication mix; and it appears to meet the user's most important information need – the need for instant access to information at a time they want it. Put all that together and you have a powerful and 'sexy' information retrieval device. The PC was never 'sexy'.

It is generally a mistake to think of mobile phone information delivery in terms of commercial online hosts or the Web. A new – or, maybe, a recycled conceptual framework is needed. To begin with we must open our minds to what really constitutes an interface – and a successful one at that. The way information is processed and packaged is surely part of the interface too. Commercial online hosts have great information but they have complicated interfaces. The WWW has attractive interfaces but there are many difficulties with the quality and quantity of the information. The turn is now that of the mobile phone – to provide a simple and familiar interface, and targeted information. That is the challenge for mobile phones – to seek out a new interface for the micro-browser crafted from information, rather than from icons, buttons, frames etc.

The Web is essentially a document based information *system* (hence the term 'pages'). The mobile phone will provide information *services*. That makes quite a difference. Why services? Services, because the phone will need context and location-sensitive information which serves information needs in the now, on the move. The WWW has become a mass electronic publication system and it is that which is the key to its enormous growth. The mobile phone will actually oversee the reduction of the information pile into messages and word-, picture- and sound-bites. Mobile information services are not a natural extension of the Web and will need new real-time targeting and focus. We really need new information services for mobile phones, carefully planned by information providers and professionals. We need services built around truly personal needs that have much stronger mechanisms for recognising those needs than the Web has ever had.

The information lynchpin in the mobile information environment will be the human operator or information *moderator.* These moderators will probably work locally – constituting an invaluable cottage industry. This is important because many enquiries will be local and social. Moderators will use computers themselves to answer enquiries and supply this data orally or textually. Some moderators will be processes that learn and predict needs. In the 'always-on' networked phone environment to come with the GPRS service to follow GSM, moderators will be able to send information on spec – relevance feedback will follow if the phone can follow you to its recommended destination!

There is another aspect to the mobile phone that is likely to capture a lot of interest – especially amongst the beleaguered commercial online hosts. It is perceived by the population at large to be a high-cost environment. This has not proved to be an impediment to its take up or use. People happily pay far more than they do for their traditional phone. This is partly to do with its business origins and partly to do with its remarkably high status. This, together with its limited ability to handle large quantities of data at speed, will mean that it is highly likely that the mobile phone information environment will be a highly competitive one in which you can get away with charging for information. It is not 'tainted' with liberal information-for-free philosophies, nor are government and education involved. This is a dogma-free and economically rich environment. Just the environment for innovation.

The mass production of information WWW style is more like a feature of an industrial society than an information society, because of the way in which quality and individuality has been sacrificed for the great God production. The Web takes little cognisance of the fact that, for instance, financial, military, and medical information domains are unquestionably built upon quality. Quality control will become a major concern for mobile information services because their users will pay for quality, not for the information itself.

The digital mobile phone offers an information service that is a much closer approximation to real life. The mobile phone represents a genuinely popular (mass) platform for seeking and finding information, not a platform that supplies data in mass. You cannot really say that of the PC, despite the giant strides it has taken towards this goal. What will surely power drive the use of the phone as an information retrieving medium will be the vast and ever-increasing amounts of real-time information becoming available. That is why you need information on the move – infomobility. Mobile phone delivered information will not only prove to be a very hot topic – and will be so for some time – but it is also something that will turn what we know on its head. We are going to have to execute some conceptual and intellectual somersaults if we are to get the most out of the latest information kid on the block. A move from virtual to human reality may prove disturbing!

Conclusion

There is a natural (orderly) tendency among us, which tends to think in terms of all communication forms and information seeking activities migrating towards one platform. That platform has been variously thought to be television, the personal computer, and now that claim has passed to the mobile phone. However, it is probably more likely that the mobile phone represents an expansion in the number of platforms, rather than a reduction or consolidation in them. As we move away from the concept of the *user* as a passive member of a featureless mass, fed information batch-processing style, to one of an *information player* (or I-player) – an active individual who performs many information roles, we should understand too that different players require different platforms.

In terms of mobile phone technology the UK probably leads the world by a couple of years. This was never the case with the Internet. Here is a golden opportunity to set the information pace. Economic success in the information and communication field rests on having a mass market for a product people are willing to pay a premium for. With the digital mobile phone you have it.

British official publishing in the Internet Age

Chris Pond[1]
Reference and Reader Services, Department of the Library, House of Commons, London
Pondcc@parliament.uk

This paper examines the state of UK official Internet publishing in 2000. It discusses issues of design, content, archiving, electronic communication, copyright and points out areas where action would be desirable.

Background

Lavinia Byrne, speaking in March 2000, used a memorable phrase to describe the advent of the Internet, and especially of free telephone access to it. Characterising it as a stupendous disruption to the established order of things, she said 'Nothing so big has happened to public communications since the Protestant Reformation' [1]. Miss Byrne, who is both a prominent Catholic and a librarian, knew what she was talking about. The Internet Age has without doubt provided a new start, as did the Reformation, sweeping away established patterns and conventions of both providing and retrieving information – as much for Government as for anyone else.

For some 23 years, the author of this article has lectured on the Aslib British Official Publications Course, speaking over the years to some 750 librarians specialising in the field. Insofar as any subject is well defined by the worry-factor – that is, the problems practitioners encounter and put questions about – the Aslib Course over that quarter century provided much food for thought.

The late 1970s were the era of grey literature – the departmental typescript 'publications' then mostly uncatalogued and unindexed, which were made possible by another technical development, the photocopier [2]. A major concern of the 1980s was the growing tendency to charge for previously published information and to make information a commodity, sellable and buyable [3]. To this was added ten years later the agitation for a reduction in the cost of Hansard and parliamentary publications [4], and the extremely high cost of the then-new technology such as CD-Rom. There were also worries that commercial organisations would take over publications from HMSO, which they would then issue at a vastly inflated price. And underlying all these questions was the universal concern, 'How do I obtain for my users a copy of the ——paper, issued yesterday?'

In the late 1970s, electronic publishing, in the form then of that proto-Internet, Prestel, became available. The question was how to exploit it. The present author wrote in 1981:

> there is no doubt that the technology exists whereby [say] the result of a division in the House of Commons could be keyed in by the officer recording it, and could be very rapidly and probably automatically formatted in a dozen different ways, made to generate the House's procedural record, the file of information to Members, give voting records to the Whips, and be pumped via Prestel to every home in the country that wanted it [5, p. 418].

This piece of crystal ball gazing has only come about during 2000, and then not fully. Prestel, of course, was undermined by a combination of high cost and limited capacity.

Some of these concerns have been addressed through the advent of the Internet, and the growing inclination of Government and Parliament to make the maximum amount of information available on the Internet, but other problems have arisen as they always do. The purpose of this article is to record some of the current issues in electronic official publishing and where necessary to indicate trends and to suggest remedies.

---

[1] The opinions expressed in this article are the author's own and do not represent any official view of the Library.

Some trends and issues

The basic form and skeleton of Internet databases

At the moment, norms and conventions that are taken for granted in print-on-paper publishing do not apply in electronic format [6]. Early books, for instance, did not always have page-numbering, tables of contents, or title pages. The index was a later invention. Among government websites, certain conventions seem to be beginning to apply: 89% of sites have a 'statement of current activities', 66% a 'what's new' section, 59% FAQs on regulatory material [7].

By analogy with the book, the home page is the title page, the search engine (even Muscat, which seems to be dreaded by many information practitioners) the index. Lacking, perhaps especially in government websites, are a contents list in any coherent or standard form, page and chapter numbering and date and place of publication. All these are taken for granted in paper publication – and the lack of them much deprecated, as in the case of grey literature – so why should they be absent or deficient in electronic publication? Many of the publications referenced in this article were referred to on the Web, but to get page numbering it was then necessary to go to hard copy.

This latter is extremely important, especially dealing with government information. We are told generally when the site has last been updated, but that is insufficient. A date of establishing a text, and of its subsequent revision, would be extremely helpful on every page, in just the same way as browsers print the date of generating the print at the foot of each sheet.

IAG Champions, which is a group of 36 senior civil servants working interdepartmentally, have drawn up guidance for government websites [8]. These include mandatory items to include, for instance, ministerial responsibilities, command papers, Bills before Parliament, aims and objectives. This is a useful start, but detailed design and navigation requirements have yet to be tackled, as has the question of archiving.

Free provision

A major advantage of the Internet has been the extension of free provision of material. If anyone had suggested in 1990 that Acts, Statutory Instruments, Select Committee reports, Hansard or the Weekly Information Bulletin would be available electronically, free of charge, he or she would quickly have been despatched to some bibliographical Bedlam. Even the campaigning authors of *What Price Hansard*? contented themselves in 1994 with recommending 'Parliament should determine to make Hansard available on line *cheaply... on the Internet at a modest charge to subscribers* [4, p. 17]' [my italics]. Yet in the space of seven years, free provision of all the items listed had been accomplished. Of course, Net publication of these documents was not achieved overnight. The influential Information Committee Report [9], which gave rise to the presentation of most parliamentary texts on the Internet, owed much to a growing realisation among parliamentary officials that democracy was not served well by making parliamentary publications difficult and expensive to obtain. International comparators, such as the USA and Australia, were pointing to free provision of such texts becoming the norm. In the UK Parliament, the disaggregation of the two Houses' printing costs and the establishment of proper contracts with HMSO/TSO helped this process enormously, because the publisher's fears that free presentation on the Net would affect paper sales were no longer relevant.

Subscription-based web services offering value-added access to official material may now actually be at a disadvantage because of the overwhelming assumption that official information should be free in the interests of democratic accessibility. This view is in contrast to one advanced by David Lennon [10] of Lexis-Nexis Europe, who argues that free information may become fools' gold. However, in the official sector the prevailing ethos has certainly been to make basic provision free as part of the process of enrichment of citizenship. It remains to be seen how subscription services rooted in this sector, such as UKOP online and BOPCAS, will fare.

Wider availability of local material

As with Parliament, so with Government. Not just central government, but local government as well. Net publishing has made available a whole raft of local government publications which, to say the least, were highly fugitive in hard-copy only days. Various estimates were made of the proportion of local government publications that found their way into the national collection: and the general figure was thought to be less than 10%. On the Net, all sorts of material are now available nationally and, of course, globally that would previously have found its way into only a few local libraries.

This is not just large councils. In the County of Essex, for instance, the County Council's website includes full text of documents such as the Replacement Structure Plan and the Essex Coastal Strategy. At district council level (and only three of the twelve district councils in Essex are without a website) full text is available of Brentwood's Best Value Community Plan, or their Press Notice on the Town Centre Working Group. At parish/town level, even, the Loughton Town Council website included agendas and minutes, procedural documents which have always been difficult to find.

When the author worked in a large public reference library in the 1970s, we maintained an extensive collection of town guides, and very difficult they were to obtain and to store (being glossy, of all shapes and sizes, and virtually never bearing an SBN or price). The equivalent information can now almost always be obtained from council websites, and (with luck) it might be up-to-date, and not five or six years old. Admittedly, it is not consistent, but where else would you look for days out in Essex or Huntingdonshire, a potted history of Kelvedon Hatch, or details of lexicographers who lived in Loughton?

Government ancillary publication

As well as primary texts, a great deal of secondary government publication is now on the Internet. Consultation documents were one of the banes of 1970s librarians, especially before the advent of the *House of Commons Weekly Information Bulletin* in 1978. Recently, over 60% of consultation documents have been presented on the Internet[2]. Interestingly, consultation documents have been a class of document that has generally been free, so this is the area in which savings do not result. It would be interesting to conduct a study investigating whether web availability of such documents increases the response rate.

Government participation in the Internet

A very useful overview of Government web participation is contained in the National Audit Office Report, *Government on the Web* [7]. The executive summary states neatly the mission statement of Government Intranet participation, 'to provide higher quality services directly to citizens in innovative ways at lower cost'. The ever-open and omnipresent nature of the Internet is stressed, together with the thought that 'many citizens may have the opportunity to conduct most of their business with Government direct' [7, p. 1].

The NAO works, of course, as does any body of auditors, from the first standpoint of cost reduction. An axiomatic phrase in the report is that the marginal cost of a user accessing a website is nil, whilst 'the marginal costs of handling letters, phone calls, or front-office visits are considerable'. The caution that poor web information erodes that advantage, from generating 'avoidable' email traffic is however a valid one, to which might be added the email traffic that is generated simply out of additional interest. The Head of the House of Commons Information Office estimates[3] that email enquiries have generated an *additional* 2,400 transactions a year to his office's total of about 100,000. So the NAO recommendation [7, p. 5] that citizens should be encouraged to use websites *rather* than calling or writing may not be effective.

---

[2] From a survey of the *Bulletin* late 1999.

[3] Information supplied to author.

The NAO report criticises government web publishing as 'patchy and relatively slow' [7, p. 3]. Some sites are not updated or kept under review. The DSS sites 'lack any coherent overall plan and are run on very small budgets...'

A more general difficulty is that government websites are a heterogeneous collection, arranged and structured differently. Publications are made available, or not, without any consistency, and seemingly at the whim of individual staff groups within departments. The CCTA has attempted to provide an 'overcoat' for departments and agencies in providing structured access to their websites, but it does not have any powers beyond the advisory for the content or internal arrangement of those sites. Even the sites of the four British Parliaments and Assemblies are radically different, even though they contain material of the same genre.

Web-only publishing

The desire of some organisations to publish only in electronic form is, however, a distinct problem to the information practitioner. The desire stems from the ineffable dual aims of being in the vanguard of IT progress and saving money. The most obvious case is of the National Assembly for Wales, which has a strict policy of electronic-only publication for parliamentary material. If you want to read what was said in the Assembly Chamber or a Committee, you can do so only on the Net. (Printout of the web pages is sent periodically to the Assembly's 'link libraries', but web pages printed out do not make much of a substitute for the well-tried structure of a printed book.) This is a state of affairs that has incensed certain sections of the information profession, notably the Standing Committee on Official Publications (SCOOP) [11], and would certainly appear to run counter to a dictum of the NAO report, which rules '....it will be very important to ensure that citizens [...] without web access are treated equally and do not become disadvantaged...' [7, p. 1]. Web publishing is not a panacea, any more than print-on-paper publishing was when a majority of the population was illiterate. Bodies that rely *only* on the Internet surely distance themselves from those they seek to serve. The existence of an information-poor underclass is something, more generally, that Government must vigorously guard against.

This problem was confronted in the Scottish Parliament, which despite a large investment in electronic systems, adopted a dual publishing policy, of web and paper issue, also with link libraries. A report on telematics in the future Scottish Parliament, issued before the legislation was on the Statute Book [12], envisaged a whole new scenario of participatitive democracy, extending to discussion forums and electronic voting. Its tangible result was a comprehensive website, but the common sense of providing hard copy properly produced was early apprehended by those who set up the Parliament.

Graphics

One result of the lack of consistency in official websites is in their immediate appearance. There must be an inherent conflict in the minds of those that devise and commission official websites. Should the body go for an elaborate, pretty website, or a boring, functional one? The mundane appearance and educative poverty of committee information on the Parliament website was recently criticised by the Liaison Committee [13]. On the other hand, whizzy graphics are all very well, but in pure information retrieval terms they are a menace, particularly if you do not have a dedicated line, know how to switch them off, or have a state of the art PC. Information practitioners seem often to single out the website of the Department of Culture, Media and Sport for criticism in this regard, but there are others. Conversely, the Department of Environment, Transport and the Regions site generally seems to receive approval – information people seem to like its 'egg' on the home page, with pictograms and text indicating the structure of the site – a worthy electronic successor to the contents page in a book.

The Central Computer and Telecommunications Agency (CCTA) who co-ordinate Government's (but not Parliament's) IT activity, launched its www.open.gov.uk website in the summer of 1994. The very choice of that URL was relevant in that it signalled the intention of the Government to use the Internet as part of its opening up of government processes. Its near contemporaries, the Citizen's Charter and Code of Practice on Access to Government Information were, arguably, part of the same process.

Within five years, by the summer of 1999, the CCTA (who are also the naming authority for gov.uk URLs) had registered 1030 domain names from 913 central, regional and local government bodies. Some 700 websites were active [14], to which have to be added the two important (UK and Scottish) Parliamentary sites, which do not (since Parliament is *not* Government), use the gov.uk suffix.

Archiving of official websites

A rather worrying NAO recommendation for the information practitioner is that organisations should ensure the whole site is well used, without dormant sections left unrevised for long periods, or 'desert' sections. There is already the temptation to cull last year's consultation document, or the 1994 policy paper; and these, if published only electronically, are likely to become not grey literature, as we had in the 70s and 80s, but the shadow literature of the future. The NAO survey indicates that only 14% of government websites had any search facility for the archive; 45% had an archive of the 'what's new' section [7, p. 13], but the report is silent on the all-important question of archiving policy for previous publications and old editions: indeed only 41% of sites had publications downloadable, and that presumably excludes any consideration of the *proportion* of publications so accessible. SCOOP has been so worried about the 'here today, gone tomorrow' aspect of the Web – to which one might add 'here today, still here tomorrow, but in an amended form' – that it arranged a seminar on archiving official web material[4].

The lack of a consistent archiving policy is a grave deterrent to abandoning paper sources altogether. Press Notices maintained a shadowy, half known existence, for some decades. The House of Commons Library, almost alone among libraries, maintains a huge collection of [paper] Press Notices from government departments back to the 60s, and needs to access them once or twice a week. It must have confidence that electronic versions will be similarly accessible in twenty or so years' time in order to abandon its long-standing practice.

Websites and electronic communication with government

Much of the NAO survey is concerned with the possibility of electronic communication with government – for instance, the filing of a VAT return or housing benefit claim electronically, thus saving a great deal of time and money. The advent of XML (Extensible Mark-up Language) technology will probably provide a great impetus to these developments. Simon Moore, writing in *Public Service Magazine* (the members' journal of the Association of First Division Civil Servants – the 'Mandarins' Union' as the press term them), sees XML as a tool for the sharing of information between government and local authority departments. Citing a computer consultant, he suggests that XML might facilitate the detection, for example, of social security fraud, and check automatically where a benefit claimant also has a parking permit for a new car, or works in one area whilst claiming housing benefit in another [15].

All this is perhaps a little far from the humdrum nature of government websites, and information practitioners' use of them, but it does indicate the impetus in government behind current IT developments. Where up till now the *provision* of information *to* the public had been paramount, more and more government Internet presence will be to *collect* data from them, and to exchange that information as it thinks necessary. The present author is not one who habitually thinks first of the 'big brother' nature of government IT, but data protection and privacy issues are surely thrown into focus by developments such as this and the Newham feature recognition project [16]. It can only be a matter of time until the City of London Police June 18 riot suspects' photographs pages [17] are linked with such technology. This site is after all only a modern technology equivalent of the nineteenth century *Police Gazette*. We can surely expect coalescing of police official web presence and what was heretofore unpublished or available in ephemeral form (posters, billboards, etc) – in just the same way as before electronic publishing.

---

[4] To be held in London on 12 May 2000, after this article was sent for press.

The Government now aims to put enough material on its websites to allow 25% of transactions to be dealt with electronically, including the publication of documents such as green and white papers.

Copyright

A concomitant of issue over the Internet is undoubtedly a loosening of copyright control: that applies as much to governments and parliaments as to companies and private individuals. When HMSO was a public department, and a trading fund, yet the Controller of HMSO was the controller also of Crown Copyright, and the administrator of Parliamentary Copyright, a conflict of interest was sometimes likely to occur, in that free presentation over the Internet was seen to be a recipe for diminishing sales and thus income to the trading fund. Separation of copyright administration and publishing ended that possible conflict of interest. Although Net publishing did not end the enforcement of official copyrights *de jure*, as suggested by the Hansard Society authors [4, p. 12], it has considerably slackened the control *de facto*. A recent author in this field has stated the situation succinctly:

> If an author puts up a web page...the whole ethos of the Web and its founders is that the information should be used to its fullest extent by copying or linking...[18]

An issue that needs resolution – oversight

What is most needed now, in the fragmented world of central government publishing, is some authoritative body to set down basic standards in areas such as construction, currency, and archiving and to ensure its departments and agencies adhere to them. For print media, a useful start was made by SCOOP in their *Code of Practice for Official Publishing* [19] but this was an attempt by a pressure group to impose order from outside. What is needed is direction from inside. The IAG Champions initiative is a step in the right direction. Perhaps a partnership between this interdepartmental working party and the information profession would be a relevant next initiative? The audit on implementation, and the setting of target dates for compliance set by IAG Champions is something the profession will have to watch closely. Archiving is something not dealt with so far by this body. It should be.

Awareness of problems like these is not new. In the government sector, the Cabinet Office/Information Technology Advisory Panel report of 1983 dealt with them fully [3, p. 36], suggesting 'such problems must be faced and overcome'. They never were: they should be now. The NAO report [7, p. 58] says rather euphemistically 'the current central machinery for co-ordinating and promoting the development of government on the Web is rather weakly developed'. They conclude, '...it is now hard for citizens to navigate central government websites' [7, p. 58]. The overwhelming need is for norms to be laid down for minimum standards for all gov.uk (and parliament.uk) websites.

For local government bodies, similar guidance is also desirable. This is much less likely to be achieved; though the Local Government Association and National Association of Local Councils might take a lead. The IAG Champions' website gives voice to a rather pious hope that local and other government organisations might take note and follow suit, but the writer has seen no evidence that they have.

The Internet is a wonderfully free and easy publishing medium, and among civil libertarians it would be seen as almost sacrilegious to tame and control it. But where information is to be offered to the citizen, surely basic rules and conventions should apply, and be made to apply. Worthy initiatives have been made in this direction: it is for the information profession to ensure they are carried forward in a suitable way.

References

1. BBC Radio 4. *Today,* 7 March 2000, 07.49 GMT, www.bbc.co.uk

2. See, for example: Jordan, G. Grey papers. *Political Quarterly,* 48(1), 1977, 30–43.

3. *Making a business of information.* London: Cabinet Office, 1984.

4. See, for example: Lester, A. *et al. What price Hansard?* London: Hansard Society, 1994.

5. Pond, C.C. Current parliamentary and government legislation, prospects and events. *Aslib Proceedings,* 33(11/12), 1981, 418–426.

6. See, for example: Meadows, J. Is the future a different country? *Refer,* 16(1), 2000, 10–13.

7. National Audit Office. *Government on the Web.* HC 87, 1999–2000, 13. Also available at: www.nao.gov/publications/nao-reports/990087es.paf. See also Jellinek, D. Official UK, the essential guide to government websites. TSO, 1998.

8. IAG Champions. www.iagchampions.gov.uk/Guidelines/websites/default.asp

9. *Electronic publication of House Documents, Information Committee 1st report.* HC 328, 1995/96.

10. Lennon, D. The future of 'free' information in the age of the Internet. *Aslib Proceedings,* 51(9), 1999, 285–289.

11. See, for example: Picton, H. Scoop News, *Refer,* 16(1), 2000, 25.

12. *A Parliament for the Millennium, report by the Advisory Committee on Telematics for the Scottish Parliament.* Edinburgh: John Wheatley Centre, n.d. [1997].

13. *1st report of the Liaison Committee, 1999–2000.* HC 300, paragraphs 100–105.

14. Statistics in this part taken, with other information, from: Crump, J. Government and the Internet. *Government IT,* February 2000, 6–8.

15. Moore, S. XML – an electronic Holy Grail? *Public Service Magazine,* 3(2), 2000, 30–32. [This subject is also addressed in: Part 5 of *The Informal Economy, a report by Lord Grabiner QC.* HM Treasury, 2000.]

16. See, for example: Underhill, W. Britain – Big Ben or big brother? *Newsweek,* CXXXV, 9, 28 Feb 2000, 31.

17. City of London Police. www.CityofLondon.gov.uk/citypolice/index.htm

18. For a detailed discussion see: Eisenschnitz, T. Internet law and Information Policy. *Aslib Proceedings,* 50(9), 1998, 267–73.

19. *Code of practice for government department or agency documents issued to the public.* Swanlow, Cheshire: SCOOP, March 1992.

Acknowledgement

I should like to thank my colleague, Janice Davies, for reading and commenting on the draft of this article.

An observational study of the information seeking behaviour of Members of Parliament in the United Kingdom

Robert Orton, Rita Marcella and Graeme Baxter
School of Information and Media, The Robert Gordon University, Garthdee Road, Aberdeen, AB10 7QE
r.marcella@rgu.ac.uk

This paper describes the results of an observational study of the information seeking behaviour of Members of Parliament. It is argued that political life functions on a flow of information and that information itself is a prime resource. The study sought to determine the characteristics that impact upon the manner in which information is sourced and used by parliamentarians. For this project a shadowing methodology was adopted. The conclusions drawn include that UK Members of Parliament are subject to a range of triggers of information seeking, in an information rich environment where the amount of information presents decision makers with difficulties. Public accountability results in demands from a wide variety of individuals, groups and associations. MPs are also expected to be knowledgeable about a wide variety of issues, while information need is frequently unpredictable and reactive. Members must be efficient information gatherers and managers and they must be flexible in their attitudes to information seeking, while retaining the capacity to critically appraise the quality of sources.

Introduction and methodology

It might be argued that representative members of legislative assemblies are the most influential decision-makers in any society. As such the quality of information feeding into and impacting upon the decision making process is of paramount importance in ensuring that informed and effective government can take place. There are rising expectations amongst constituents for a speedy and informed response from Members. Equally, the abundance of information, often unsolicited, available to MPs does not necessarily improve the quality of its retrieval, evaluation and use.

Despite the significance of the topic, there has been little research into MPs' information seeking behaviour. Barker and Rush [1] noted that:

> politicians are advocates of public issues and use information of various kinds to support their opinions... This basic device of choosing information on an issue to suit an already firmly settled opinion appeared in our survey interviews in various forms.

Shepherd [2] provided the MP's perspective on information need, highlighting Members' 'incessant and relentless need for information' as one of the characteristics of the role. Factors identified by Shepherd as affecting information need include: the high level of demand from constituents, the growth in the lobbying phenomenon, the need for a speedy response and the continuing growth in need amongst MPs for digested and easily retrievable information. Shepherd's paper forms part of a useful collection of essays, edited by Englefield, which also includes discussions of the role of Members' personal staff, by Griffin-Smith [3] and Colquhoun [4], and special aspects of information need, such as Clements' [5] description of the use of statistics by politicians. Franklin and Norton [6] analysed the special function of the parliamentary question as an information seeking tool and detailed the results of a survey into its use.

Levin and Stonefield [7] found that the House of Commons Library was considered to be a crucial source of information in support of decision making, by both MPs and their staff; while Tanfield [8] examined the challenges facing the House of Commons Library Department in seeking to support the differing information needs of Members and the public. Lock [9] described the role of subject specialists as part of the Library's services. Two earlier MSc dissertations focused upon the House of Commons Library service to MPs [10] and the information needs of Members of the European Parliament [11]. Martell [12] described the services of the Parliamentary Labour Party Resource Centre.

The annual House of Commons Commission report includes an annex [13] appraising the performance of the Library. In 1998/99, for example, the report included an acknowledgement that growth in demand for Library services may 'have begun to slow down ... partly attributed to the Library's continuing efforts to provide information to Members, and their staff, in electronic form'. The report also referred to the results of a recent internal survey showing that the Library continued to be highly regarded by users.

Englefield [14], Menhennet and Wainwright [15], Siswell [16] and Wainwright [17] discussed the design and effectiveness of POLIS, the Parliamentary Online Information System; while Hancock-Beaulieu [18] investigated the use of online searching at the House of Commons. Michalowski and Nawrocka [19] described the developments in and growing demands on parliamentary research services through a comparative study of eleven parliaments.

This paper describes a study that sought to determine the unique characteristics of, and special factors impacting upon, the information seeking behaviour of Members of Parliament in the United Kingdom.

The objectives were:

- to gather data about the sources of information available to MPs;
- to identify strengths and weaknesses in observed approaches to information seeking; and
- to consider which environmental and user characteristics impacted upon information seeking behaviour.

An observational technique was used in order to gain an in-depth understanding of the information seeking behaviour of Members of Parliament, as it was felt that other data collection tools, such as questionnaires, might gather very superficial responses and might be completed by staff members rather than MPs. While it is fully acknowledged that the small sample used in the present study (two MPs from the same party) is not generalisable to all MPs, it is felt that the findings may help to suggest hypotheses that could be tested more generally in the future. The observational approach was considered appropriate as 'observation is a good technique for looking at user behaviour because it is objective and only records what actually happened' [20].

For this study two Members of Parliament (one for an English constituency, the other representing a Scottish seat) were selected, both Labour Party backbenchers, each having entered Parliament for the first time in 1997. The researcher observed the two cases over an extended period of four weeks, and a log was kept detailing each Member's information seeking behaviour. Direct information seeking by the cases was recorded, as were instances where the Members brought about information seeking through others. The log was analysed by classifying activities and events. Data gathered by the researcher allowed hypotheses to emerge with respect to the information handling competencies required of a Member of Parliament.

The Members observed for the present study

Member 1, aged 44, was formerly a schoolteacher and is active in the Fabian Society and in the Co-operative Party. The Member's political interests include the environment and s/he currently sits on the Environmental Audit Committee. S/he has been instrumental in raising the profile of many issues, including the Home Zones Campaign, which relates to the slowing down of motor cars in designated streets and in the use of alternative fuels such as Liquid Petroleum Gas. An active 'Westminster' politician, the Member headed part of the campaign to elect Tony Blair as

Labour Party leader. S/he claims to have received more media coverage than any other new MP.

Member 2, aged 48, was previously an advocate prior to the election and has a long record of Labour Party involvement, including the post of Permanent Private Secretary to Donald Dewar. The Member sat on the Select Committee on Northern Irish Affairs and was an official observer at the South African elections. S/he currently sits on the Select Committee on Public Administration drafting the Freedom of Information Bill.

At Westminster, MPs are restricted to a maximum of three staff. The experience and qualifications of a Member's staff have a bearing on the quality of information seeking undertaken. Many are young and inexperienced. Most interns are graduates and may have research experience.

The workload of MPs is perceived to have increased. Today MPs tend to receive larger quantities than previously of post directly from constituents covering a wide range of subjects and issues, although many of the topics raised do not lie within the purview of the Members. A large volume of mail is also received from increasingly professional pressure groups.

The offices observed for this study had differing staff profiles. Three members of staff were employed at the Case 1 Westminster office on a part-time basis. The most senior post was that of political advisor, a doctorate holder with many years of Westminster experience for several MPs, whose duties included preparation of briefs, speechwriting, compiling parliamentary questions and research. A secretary was responsible largely for clerical and administrative duties, including organising the MP's diary, arranging meetings and manning the telephones. Collectively these two members of staff had almost fifty years of Westminster experience. A research volunteer was also employed to prepare briefs, research and prepare parliamentary questions and liaise with media and other groups. Member 1's constituency office was staffed on a part-time basis, providing constituency secretarial and administrative support, organising the Member's constituency diary and the weekly surgery.

Two members of staff were employed at the Case 2 Westminster office, one full-time, one on a part-time basis. The full-time assistant had previously worked at the House of Commons Library and was responsible for office administration, preparation of press releases and research. A part-time research assistant (shared with a Minister) provided research, prepared press releases and briefed the MP on a wide range of topics. Member 2's constituency office was staffed by three employees: a full-time secretary, who had worked for the former constituency MP, a part-time administrator and a young, part-time clerical assistant.

In the Case 1 office, staff meetings were not held and staff were not systematically apprised of information need. This style had drawbacks in that the strict demarcation of roles within the office may have hindered the flow of information. The office often felt hectic and pressurised and the lack of full-time staff was not conducive to the full exploration of information possibilities.

The Case 2 office was characterised by its openness and by the manner in which the work in hand was discussed. Regular staff meetings were held which allowed for free discussion and delegation of the workload. Staff could develop a clear understanding of the information required and of the parameters of the search to be conducted.

Case MPs' use of information services

The House of Commons Library:

> provides impartial and confidential information and research services for Members of Parliament in connection with their parliamentary duties, and for staff working on behalf of Members [21].

The Research Service at the Derby Gate Library is staffed by subject specialists who prepare research papers on topical issues, which may be requested by Ministers or their staff. Members who wish to find information on a particular subject often start their search by determining whether such a paper has been produced. During observation, the Library's Social and General Statistics Section was asked by Case 2 to provide information on the Scottish and Welsh elections to their devolved Houses. Members frequently seek information relating to their constituencies. The Library produces these to order, giving summaries of employment, unemployment, earnings and other information. Constituency profile requests are handled by the Statistics Resources Unit within a few days.

The Members' Library in the Palace of Westminster stocks United Kingdom official publications, a wide range of reference materials, numerous newspapers, and political, economic and legal journals. MPs can call in person or make contact with the Library's enquiry service by letter, email or telephone. The staff employed at the Members' Library also act as intermediaries between MPs and the Research Service staff. The more complex the enquiry the longer the time taken to deliver the information. For this reason Members are encouraged to be specific as to their requirements. Michalowski and Nawrocka note that 'requests for private briefings by researchers for individual Members makes up a growing share of their [the Library's] services' [22].

Two typical instances of library use were observed:

- A phone call was received which the Member took personally. The information requested was a contact address and phone number for a specific Member of the European Parliament (MEP). The MP used the Members' helpline service and the information was received after a few minutes.
- The constituency office called requesting information on behalf of a constituent who sought information and contact details on the Social Security Ombudsman. A member of staff contacted the Members' helpline and explained the query. The information requested was ready to be collected later that day. Along with information about the specific ombudsman, library staff had included additional general information about the role and power of ombudsman services.

Member 1 used the Library five times in total during the month of observation, with staff members using library services by telephone or in person several times a week. Member 2 used the Library three times and was observed to call the Members' helpline on five occasions. One of Member 2's staff had previously been employed at the Library and it was felt that this contributed to the frequency of use of the Library, as she was familiar with the facilities and resources available.

A frequently overlooked aspect of the Library is that it provides a place where Members can work undisturbed. It is also situated close to the Members' lobby, convenient for MPs entering or leaving the Chamber. The Library's quiet room is also used by Members to write speeches. Case 1 was observed using the Library on two occasions for speechwriting, while Case 2 was logged using it once to construct a reply to a local newspaper.

Both MPs and their staff demonstrated positive attitudes to and high levels of awareness of the Library. Library services were exploited on all appropriate occasions. Neither Member used the private briefing service during observation: however, one had done so shortly before observation began.

The Parliamentary Data and Video Network (PDVN) gives access to the Internet and the parliamentary intranet; to Members and their staff in Westminster and constituency offices. The POLIS database, available via the PDVN covers a wide range of relevant material, including: parliamentary proceedings; debates; questions; legislation; reports of Members' contributions; recently deposited papers and reports; the progress of current bills; European legislative proposals; and European Union publications. The system is also linked into the Early Day Motion database and the House of Commons Library's online catalogue.

Case 1 did not personally use POLIS or the PDVN during the period of observation, although s/he was aware of their potential value. While his/her staff accessed the *Early Day Motion* database twice, this Member did not encourage staff to use electronic resources for information gathering and, despite expressing a desire to have a web page, did not see the development as a high priority. Case 2 displayed greater enthusiasm for electronic information resources and actively encouraged electronic retrieval by staff, with the MP regularly using email, the parliamentary intranet (on two observed occasions) and electronic news services via the Internet. Material was also sourced electronically on several occasions by Case 2 staff.

Each office at Westminster has a television set giving access to the parliamentary annunciator system, which provides textual information and live broadcast of debates. Televisions are also provided in corridors, committee rooms and other places. Time can be saved by Members remotely viewing debates, although this may lead to a fall in the standard of debates and a lesser likelihood of Members making a spontaneous contribution [23].

All political parties devote much time and effort to the collection and dissemination of information, in order to keep Members up to date with party initiatives and policy developments. The Parliamentary Labour Party Resource Centre is an additional party supported facility, established in 1991 and located in Westminster Hall. It provides an information and current awareness service to Labour Party Members and their staff, via weekly and daily briefing sheets, faxed to Members' offices at Westminster and in constituencies. Other useful services provided by the Centre include factsheets in support of campaigning, assistance in the preparation of press releases and newsletters for use in local constituency work.

Despite its accessibility, neither of the MPs observed used the Centre personally during the observation period. It was, however, used five times by Case 1 staff: twice for information to be used in press releases, on foxhunting and transport respectively; for information for a speech in the Chamber on transport; to collect a budget briefing pack; to gather information on the 'right to roam' legislation; and to collect a media information pack giving information on press release structure and content. Case 2 staff did not use the Centre during the observation period but had used it in the past and expected to do so in the future.

A weekly running order of House business is produced by the party Whip's office. This informs Members of the name, nature and times of debates and of their whip status, enabling MPs' diaries to be organised around the times when they must be in attendance in the House.

The Labour Party Policy Unit also disseminates policy information to Labour Party MPs, including information relating to their constituencies. The Policy Unit hosts the Excalibur database, which provides analyses of constituency information. The researcher observed one example of the Unit's services being used during the study when Member 1 requested information for a debate on public and private transport use in the case constituency.

At Westminster, MPs can meet with Ministers, officially or unofficially, and much information can be passed during such sessions. A Member's ability to obtain a ministerial hearing depends on timing, on the matter(s) to be discussed and on the particular MP's status within the party. During the observation period, Case 1 met with a Minister on three occasions, twice at Westminster and once during a constituency visit, discussing respectively, transport, health and the opening of a new facility at a local college. Case 2 had one meeting with a Parliamentary Under Secretary of State, regarding social security payments for ancillary staff in schools.

Increasingly 'special advisors' are used by government. Through such advisors, MPs have access to high quality information. Neither case approached special advisors during observation.

Away from Westminster, Members have their constituency parties who also seek to influence the Member's actions. An MP with few local connections may be disadvantaged through having little contact with their local party or the networks that extend from it. Members who have represented their constituencies for a lengthy period have had the opportunity to build relationships, which add layers of information possibilities through personal contact and knowledge. Equally, 'constituency' MPs often have a far greater involvement with their local parties than 'Westminster' MPs, as a result of which local information may be more readily available to them. Constituency visits also provide opportunities to open or strengthen local information channels.

Involvement in parliamentary groups provides Members with opportunities to impact upon government policy, to raise the profile of issues and to work together with others who have similar interests and concerns. Such groups have a huge range of specialist or technical information available to them through their contact with special interest groups and professional associations. Parliamentary groups frequently arrange seminars and conferences to review issues and encourage debate.

One of the observed MPs facilitated information flow for a parliamentary group, the All Party Parliamentary Inland Waterways Group. Literature received from the Inland Waterways Authority was distributed by the MP to others in the group. The new information encouraged the group to step up its promotional activities.

Lobbying, or interest articulation, has become much more significant and professional in recent decades. Lobbying may take a number of forms, including petitions, organised writing campaigns to MPs and direct contact between interest groups and MPs. Lobbying can be undertaken by commercial or non-commercial groups, by individuals or professional lobbying firms. Many of these groups maintain offices at or near Westminster, from which they liaise with MPs, build contacts and disseminate information.

Both cases maintained links with lobby groups. The bottling plant of the largest whisky distiller in Scotland is located in Member 2's constituency and this Member has spoken in debate in the House on behalf of the whisky industry. The researcher observed one meeting with representatives of the whisky industry at Westminster. A regular supply of information comes to the Member from the Scottish Whisky Association.

Member 1 has an interest in environmental policy and is actively involved with environmental groups promoting 'clean' fuels, with companies producing such fuels and with the lobbying organisations employed on their behalf. With such an issue, scientific and technical information is required and this is frequently disseminated by interest groups or the lobbyists representing their interests. Information produced in this way is not designed to be objective and the case MP did not rely solely on sources provided in support of one side of the argument, but sought to gain a balanced perspective. Lobby groups can often provide unique and valuable information, but are likely to be subjective and partisan sources.

Relations with national and local media may be regarded as a two-way conduit, with information flowing in both directions. Both case MPs have columns in their local newspapers and both appear regularly on local radio programmes. The researcher logged an average of two to three appearances per week on local radio by the case MPs, and observed the negotiations between the MPs and radio reporters about topical, newsworthy items. Local media can monitor and distil a large body of public opinion with a far wider reach than any political party. Case 1 sought the opinion of a media contact on the level of enthusiasm in the local community for a local college seeking university status.

The parliamentary press lobby at Westminster contains representatives of national, regional and local media and is an integral part of the life of the House, used by (and using) Members as information sources. Given that the more senior of the Westminster press corps have a highly developed intelligence network and a wealth of experience of the House, their usefulness to MPs is recognised. The knowledge network displays strong links, the press and MPs mixing in a very small circuit. The observed Members both have media links developed throughout their political careers. One of the cases has particularly significant media connections including personal friendship with the political editor of a national daily. MPs and the media tend to meet and share information informally in the refreshment rooms at Westminster.

Both MPs conducted regular media watch sessions, where they and their staff scan the local and national press for pertinent items. Newspaper stories that could form the basis of a future campaign were particularly sought by one of the MPs, as were items relating to the Freedom of Information Act. The monitoring could be a brief process lasting at its shortest a few minutes or on occasion several hours if pertinent stories were found which required further research.

Information gathering techniques by Members of Parliament with respect to specific caseload material

Several logged examples illustrate the extended process of information seeking from problem identification through to resolution and a sample of these is detailed below.

Case History 1: the Child Support Agency

The first case concerned a constituent with a long running complaint against the Child Support Agency (CSA). Following the receipt of a letter from the constituent, the MP contacted the constituency office so that details could be checked. The MP then outlined the case to the staff at Westminster, one of whom was asked to read through CSA literature with the MP. The member of staff continued alone, as the MP returned to the telephone to speak to the constituent to clarify points. The MP then phoned the CSA Parliamentary Officer who was given the details of the case, with more complete details being faxed simultaneously. The resultant information led the MP to consult the CSA literature once more with a member of staff who had now produced notes on the literature. The next day the MP had an unobserved meeting with the appropriate Minister of State at which the case was raised. It was felt that the information need had been met satisfactorily at this point.

Case History 2: school ancillary workers

The MP had been contacted by two constituents about social security payments to school ancillary workers out of term time. Details of contracts, wages and correspondence had been collected into a file by constituency office staff. The MP contacted the local trade union representative by telephone, and then called the Parliamentary Officer of the Department of Social Security. The Parliamentary Officer of the relevant trade union was then contacted by the Member: this officer was able to furnish the MP with advice on specific similar difficulties and faxed relevant information to the Member's office. After considering the information gained from these sources, the Member made an appointment to meet with the Parliamentary Under Secretary of State for Social Security. The unobserved meeting was held three days later. The topic remained an ongoing area of information need, aspects of which required further clarification.

Case History 3: control and restraint techniques

Following a violent incident when a technique for controlling and restraining had been used, an Afro-Caribbean constituent had died whilst in the care of a nursing home which was situated in another constituency. The MP was asked to investigate the case by a family member. A series of telephone calls between the Member and the relative took place, followed by a meeting in the constituency office. The Member was then contacted by the local media who were interested in the story. The MP contacted the Member of Parliament in whose constituency the clinic was situated in order to gain local knowledge and develop a joint strategy. An Early Day Motion (EDM) was tabled by the case MP and a circular letter was sent and telephone calls made in order to encourage support for the EDM. As a result of the considerable publicity generated, the Parliamentary Office of the mental health group MIND contacted the case Member offering to contribute information about the conditions of patients in mental health institutions. A second organisation, INQUEST, a charitable organisation that seeks to ensure that inquests are held where disputes over cause of death have arisen, also contacted the case MP.

Together the MPs, MIND, INQUEST, the family of the deceased and the media all contributed information. Both of the MPs asked questions in the Chamber and the case Member asked a question at Prime Minister's question time, which again raised the profile of the issue. The case MP had been informed by the Speaker's office that the name of the deceased constituent could not be used in questioning as the matter was sub judice: however, interestingly, an unsolicited call was received from yet another Member at this point informing the case MP that the constituent could be named.

A number of types of information were utilised, including: statistical data on deaths in mental health institutions; guidelines and regulations regarding the notification of relatives and the support and guidance offered to relatives after a death in a mental health institution; an oral reply by a Minister on the support and guidance offered to relatives; documentary material on the systems required to be in place in mental health institutions to deal with medical emergencies resulting from the use of control and restraint techniques and drugs; a written reply from a Minister regarding staffing levels and policies for responding to emergencies; a written response from a Minister on the collection of statistical data; and literature provided by the specialist mental health and inquest organisations.

Case History 4: the Westminster Parliament and the 'West Lothian Question'

In response to a speech by William Hague raising the issue of Scottish Members having voting rights on English matters at Westminster, while English MPs do not have the same rights on matters devolved to the Scottish Parliament and Welsh Assembly, one of the case MPs sought to develop a political line on the issue. The Member began by calling together staff for a free discussion around the subject. A press release was drawn and circulated to local media, and a series of informal meetings was organised between the Member and other Scottish MPs and Members of the Scottish Parliament (MSPs).

A staff member was delegated to visit the House of Commons Library to research the question. The Parliamentary Data and Video Network (PDVN) was used to access POLIS for references to the 'West Lothian Question'. Newspaper articles dealing with the topic were also located via the PDVN. Journal articles were located in the Library and included in the materials assembled, as were the Hansard debates, second reading on Scottish devolution (1978 and 1997). The MP took the assembled research findings to work on during the recess in order to plan a strategy for the next session of Parliament.

Case History 5: the Freedom of Information Act

One of the observed MPs sat on the Select Committee on Public Administration deliberating the pre-legislative stage of the Freedom of Information Act. The quantity of information available to members of Select Committees is vast, as the consultation exercise tends to be wide ranging. This particular piece of legislation has also attracted much press coverage. The MP delegated a researcher to become familiar with the draft legislation, the consultation exercise, documentary material circulated to committee members and media coverage of the issue, in order to prepare a brief and suggest questions for the MP to ask of expert witnesses.

The researcher observed one meeting of the Select Committee, at which a Minister was questioned about the Government's proposals. Following the meeting, the Member discussed the conduct of the meeting, expressing the view that the Minister's response had been 'political' and that the information content of this particular committee session had, therefore, to be decoded 'politically' with the intentions of the various actors being understood in this light.

The MP met with another committee member informally, when the issue of the Freedom of Information Act was discussed. Other informal meetings took place between the case Member and an MP not involved with the Select Committee but of high standing in the legal profession, where the FOI Act was raised.

Conclusions

The research facilities at the House are excellent and a wide range of information sources were used to satisfy information need in the above cases. Formal and informal contacts were the most significant sources in Case Histories 1 and 2. Specialist organisations were very important sources in Case History 3 as, with their expert knowledge of the subject, they were able to provide a wealth of pertinent and usable information. Publicity generated interest amongst a wider group, which also resulted in additional information being unearthed. In Case History 4 the Library was particularly helpful as its stock of readily accessible information enabled research materials to be assembled swiftly. Case History 5 demonstrated the procedures of a Select Committee and the means employed to gather information in support of committee work.

In all cases, formal and informal sources provided pertinent information and hard copy sources were supplemented by information accessed from online services. However, there are variations in the extent to which Members will exploit electronic sources. Equally, there are likely to be differences in the extent to which Members can draw upon informal networks, both at Westminster and in the constituency. The cases of information need described above tend to demonstrate the significance of informal networks at Westminster and its 'village' atmosphere. They also show that information seeking by Members is frequently not a self-activated nor sole activity: Members are often passive recipients of information which they have not solicited and which may come to them by chance or as a result of the activities of others, who may have quite different reasons for becoming involved.

MPs used a wide range of information sources, in each instance on a relatively infrequent basis. These cases demonstrate the importance of flexibility in approach to information seeking, as well as the need for MPs and/or their staff to filter effectively the information available, in order to reach both a usable mass and to focus upon reliable and valid sources. MPs frequently require guidance from expert sources in dealing with what may be unfamiliar and highly demanding topics. Literature alone was often insufficient to explicate highly complex and often contentious issues. MPs tend to access information typically through one or more levels of intermediary, through for example research assistants and library staff. This may lead to problems in that the greater the number of mediations, the more likely it is that there will be some divergence from original information need. Thus, while MPs are subject to the same difficulties as other decision makers in using information, there are certain unique characteristics which apply to MPs and not to other professions in information rich environments.

Several conclusions have been reached as to the qualities of MPs seeking and using information, largely in agreement with those of Shepherd [2]. Firstly it must be noted that not all MPs come from an equal professional level. Personal skills and professional abilities appear to have an impact upon the way that a Member conducts his or her business. It would be fair to surmise that previous experience of information rich environments does have a bearing on a Member's ability to not only cope, but to thrive in an environment where a wide range of issues must be understood quickly and in a degree of depth. Issues arise very quickly in the world of politics and Members must be able to cope with the demands made upon them by the information possibilities and be able to retain useful information – a skill in itself.

Members of Parliament often portray themselves as being independent and autonomous actors in the political world. This assumption is a factor that must impact upon their information sourcing and handling behaviour. However, MPs who belong to an established party are not independent: the working lives of MPs represent, therefore, a collaborative endeavour. Ideology, or the need to keep in step with party lines, can often limit information seeking behaviour to seeking a reinforcement of that which they already believe, or which the party line expects them to follow. This has been noted by Barker and Rush [1, p. 15] in their study where they noted the 'device of choosing information to suit an already firmly settled opinion'. This can be viewed as a strength in what is often a hectic and rushed environment where information must be quickly collected and acted upon. It may also be deemed a weakness in that such a closed mindset limits the range of information that a Member will seek or consider in reaching a decision. Such tendencies would appear to be context dependent from the results of the present study.

The researcher did not observe any form of information audit being conducted by the central party apparatus. An audit could identify the strengths and weaknesses of information sourcing and use and help to promote best practice. Such a model would be of benefit to the party, the representative and the represented. A difficulty here is that such an initiative by the central party apparatus could be seen as jeopardising the independence and autonomy of MPs by placing Members under greater scrutiny. The differences between back and frontbenchers in terms of resources is great as frontbenchers can also draw upon ministerial departments to satisfy their information requirements. The quality of a Member's staff is also highly significant. Arrangements for staffing differ enormously, with no systematic career structure and a loss of continuity is seen as staff tend not to stay with the same Member for long. The desire to have a career structure and some form of grading system was expressed through several interviews conducted as part of this study. The difficulties of putting in place such a structure within the architecture of the central party apparatus are apparent from the failure to agree such arrangements for the Scottish Parliament, where such an initiative was rejected as impinging upon an MP's autonomy.

The triggers for information seeking amongst MPs are, in many instances, unpredictable, emanating frequently from constituents' demands and from issues receiving attention in the media. The MP is almost unique professionally in having little control over the subject areas in which s/he may have to become involved and quickly expert, in response to party, constituency, public and media demands. Information need is, therefore, very often reactive and, as a result, information seeking may be rushed, unsystematic and uncritical.

There is no other professional equivalent where a body of individuals, from diverse backgrounds and with widely varying qualifications, is expected to carry out demanding new duties with no professional or educational preparation specifically for the tasks involved.

References

1. Barker, A. and Rush, M. *The Member of Parliament and his information*. London: Allen and Unwin Ltd., 1970.

2. Shepherd, C. The Members' need for information. In: Englefield, D., ed. *Workings of Westminster.* Aldershot: Dartmouth, 1991, 25–30.

3. Griffin-Smith, J. Members' personal staff: the secretary. In: Englefield, D., ed. *Workings of Westminster.* Aldershot: Dartmouth, 1991, 31–35.

4. Colquhoun, M. Members' personal staff: the research assistant. In: Englefield, D., ed. *Workings of Westminster.* Aldershot: Dartmouth, 1991, 36–40.

5. Clements, R. Parliamentarians and figures. In: Englefield, D., ed. *Workings of Westminster.* Aldershot: Dartmouth, 1991, 121–132

6. Franklin, M. and Norton, P., eds. *Parliamentary questions.* Oxford: Clarendon Press, 1993.

7. Levin, J. and Stonefield, E. *House of Commons Library user survey: report on research amongst Members and staff.* London: Levin Associates, 1995. [Unpublished internal report]

8. Tanfield, J. *Services for Members: services for the public – the dual mandate of the Library Department of the House of Commons. Proceedings of the 59th Council and Conference, Barcelona, Spain, 22–28 August 1993, Booklet 1, Division of Research Libraries.* The Hague: IFLA, 1993.

9. Lock, G. Subject specialists in the House of Commons Library. In: Englefield, D., ed. *Workings of Westminster.* Aldershot: Dartmouth, 1991, 87–98

10. Serema, B.C. Matching MPs' information needs and information services in the House of Commons. *Library Management*, 20(3), 1999, 179–189.

11. Carcary, I., Marcella, R. and Baxter, G. The information needs of United Kingdom Members of the European Parliament. *Library Management,* 20(3), 1999, 168–178.

12. Martell, V. Providing an information service to MPs: the Labour Party Resource Centre at the House of Commons. *Assignation,* 11(2), 1994, 30–33.

13. House of Commons Commission. *Twenty first annual report 1998/99: Annex A – Department of the Library.* http://wwwpublications.parliamen.../cmselect/cmcomm/721/annex-a.4.htm (visited February 2000).

14. Englefield, D. How the parliamentary on-line information system at Westminster was planned. *INSPEL,* 16, 1982, 154–165.

15. Menhennet, D. and Wainwright, J. POLIS in Parliament: computer-based information retrieval in the House of Commons Library. *Journal of Documentation,* 38(2), 1982, 72–93.

16. Siswell, A. The parliamentary on-line information system. *The Law Librarian,* 17(1), 1986, 23–26.

17. Wainwright, J. Parliament's papers and proceedings: the use of computers. In: Englefield, D., ed. *Workings of Westminster.* Aldershot: Dartmouth, 1991, 109–120

18. Hancock-Beaulieu, M. Online searching at the House of Commons. *Library and Information Research News,* 13(49), 1990, 6–7.

19. Michalowski, J. and Nawrocka, E. New developments in parliamentary research services and technology. *INSPEL,* 33(1), 1999, 20–28.

20. Eager, C. and Oppenheim, C. An observational method for undertaking user needs studies. *Journal of Librarianship and Information Studies,* 28(11), 1996, 15–23.

21. House of Commons Library Department. *A guide to the services of the House of Commons Library.* London: House of Commons Library Department, 1999.

22. Michalowski, J. and Nawrocka, E. New developments in parliamentary research services and technology. *INSPEL,* 33(1), 1999, 23.

23. Roth, A. The lobby's 'dying gasps'. *British Journalism Review,* 10(3), 1999, 21–25.

A user-centred approach to the evaluation of Subject Based Information Gateways: case study SOSIG

Maria Monopoli and David Nicholas
Internet Studies Research Group, Department of Information Science, City University, Northampton Square, London, EC1V 0HB
m.monopoli@city.ac.uk

There has been a massive increase in the amount of information available to people as a result of the Internet and information seekers are expected to be in the position to identify and evaluate this information according to their needs. As a response to this, Subject Based Information Gateways have been introduced providing people with an organised collection of digital information. An evaluation of the Social Science Information Gateway (SOSIG), conducted by an online questionnaire, examines such issues as who the SOSIG users are, how frequently they use the service and what their reasons for using the service are.

Introduction

The proliferation of information available on the Internet has raised a number of issues concerning the quality of information and access to it. People are exposed to an ever-increasing amount of networked information and are expected, often unrealistically, to be able to handle this themselves. In many respects the Internet is like a huge vandalised library. Someone has destroyed the catalogue and removed the front matter, indexes, etc., from hundreds of thousands of books and torn and scattered what remains [1]. One response to this – and a seemingly successful one – is the development of Subject Based Information Gateways (SBIGs).

> Subject gateways are online services and sites that provide searchable and browseable catalogues of Internet based resources. Subject gateways will typically focus on a related set of academic subject areas [2].

According to Wierenga [3] 'SBIGs aim to solve the problems, which currently hinder the use of the World Wide Web'. A good example of a SBIG is the Social Science Information Gateway (SOSIG). SOSIG is an online catalogue of Internet resources on subjects related to the social sciences, such as education, philosophy, and politics. Picking-up on the analogy made earlier, it offers a non-vandalised well-trusted library [1]. Its purpose is to offer a quick and easy way of finding quality networked information – including electronic journals, digitised books, reports and papers, bibliographies, educational software, electronic newsletters, mailing lists, and home pages of key social science organisations, that can support users in their work and study. The Gateway is the Internet equivalent of an academic library, where information and computer scientists use a combination of library applications and practices and database technology in order to locate, assess and describe networked resources from all around the world, with an emphasis especially on European resources. Users can either search or browse these resources according to their needs.

Aims and objectives

This article provides the result of an evaluation of the use of the SOSIG gateway. The evaluation was conducted on behalf of the Institute for Learning and Research Technology based in Bristol, which also hosts SOSIG. Its purpose was really to determine what users thought of the service. The question we posed was how do people use the SOSIG gateway for obtaining information?

Inevitably, as a result of asking SOSIG users this question, we would also be obtaining their views on a number of issues concerning remote access to the contents and services of an information system based on the Internet. A comprehensive scan of the literature had disclosed that there was very little in the way of user evaluation of online subject gateways. This study it was hoped would add to our knowledge in this area.

Plainly, digital library designers, like those involved with SOSIG, require use data. For an information system to be truly useful and valuable to users designers need first to understand the context of the person's work, his or her information acts, such as information searching. Dix *et al.* [4] points out that even if one has the best methodology and model in the design of an information system, one still needs to evaluate this system in order to satisfy users' information expectations. Because SBIGs are a relatively new type of information service this meant that, to some extent, user requirements are still really unknown. Their Internet base also implies that they suffer from the remoteness of the developers from their users.

Literature review

The fact that electronic resources, such as electronic journals and websites, have become a major source of information for many people has highlighted the need to know more about what people think about them. Results of user studies have indicated that people are well favoured to the general concept of electronic resources. Users appear willing to use a combination of paper and electronic information, with a preference for electronic where available [5]. They increasingly want to be able to work from home and to access electronic resources from there, thus keeping up to date with their disciplines, saving time, and managing electronic literature more easily and efficiently. Ideally, they would like to get the full-text of an electronic document from the first source they use. In particular they do not enjoy searching an electronic resource to find a bibliographic reference and then having to use another system to locate the full text of this item [6]. However, electronic resources are mainly used to supplement – but not to replace – traditional modes of communication, such as a traditional library [7]. This can be explained by the fact that users prefer familiar and well-established information resources. They are reluctant to attempt new information searching practices until they are totally convinced of their efficacy and they feel comfortable with them [8, 9, 10, 11].

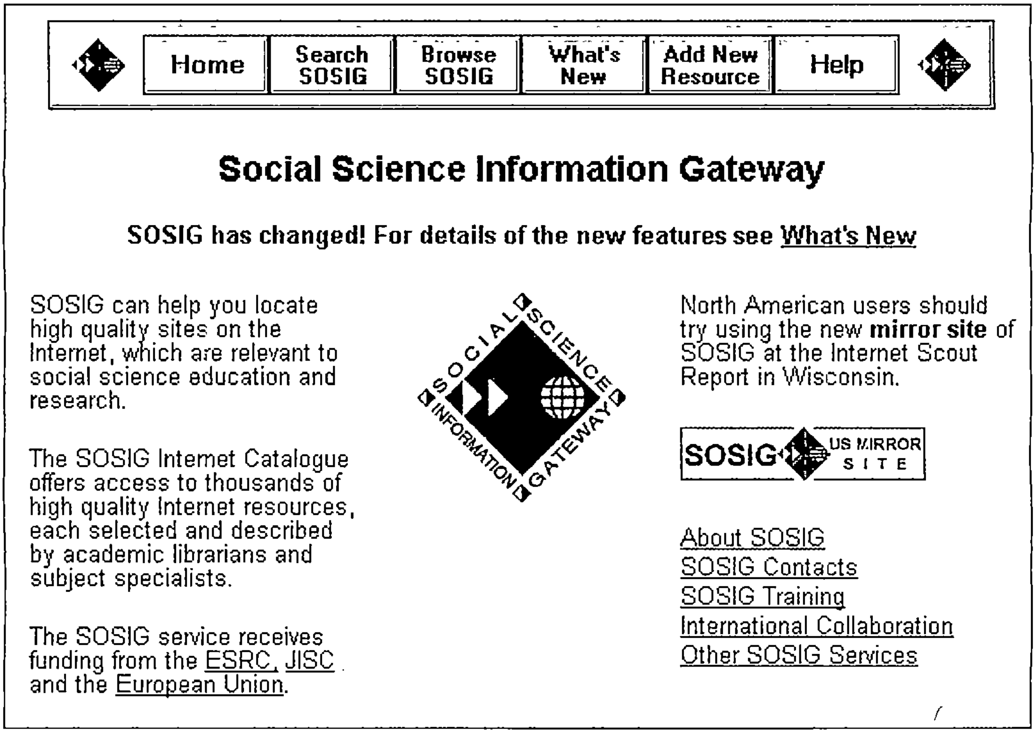
The other doubts that have been raised are about the accessibility of electronic information. Users want fast and easy access to information without being required to memorise usernames and passwords in order to use an online information system. If users experience significant barriers in registration and log-in procedures that are time-consuming or difficult to complete, they might abandon their attempts to use an information system [12]. The next major important requirement is to be able to browse, search information stored by an information system and easily print or download data that they are interested in [13].

An investigation of the use of automated chemical information systems by academic chemists in the UK [14, 15] concluded with ten recommendations to promote the use of electronic information by academic chemists. These included raising awareness of available resources, creating resources that are easy to use and reasonably priced, the provision of training and technical support, and raising levels of computer literacy.

SOSIG

Developed in 1994 the SOSIG gateway is an ESRC and JISC funded project. SOSIG provides users with a number of services including online help, browse and search facilities, a mailing list, a thesaurus, and an opportunity to suggest new resources. On 25 February 2000 SOSIG designers launched a redesigned and upgraded service (Figure 1). The new services include: 'my account', which enables users to create their own, tailored SOSIG information services and the 'grapevine', a source of research, training and development opportunities in the social sciences. The evaluation we present here, however, is based on

Figure 1: *SOSIG Home page*



Social Science Information Gateway

SOSIG has changed! For details of the new features see What's New

SOSIG can help you locate high quality sites on the Internet, which are relevant to social science education and research.

The SOSIG Internet Catalogue offers access to thousands of high quality Internet resources, each selected and described by academic librarians and subject specialists.

The SOSIG service receives funding from the ESRC, JISC and the European Union.

North American users should try using the new **mirror site** of SOSIG at the Internet Scout Report in Wisconsin.

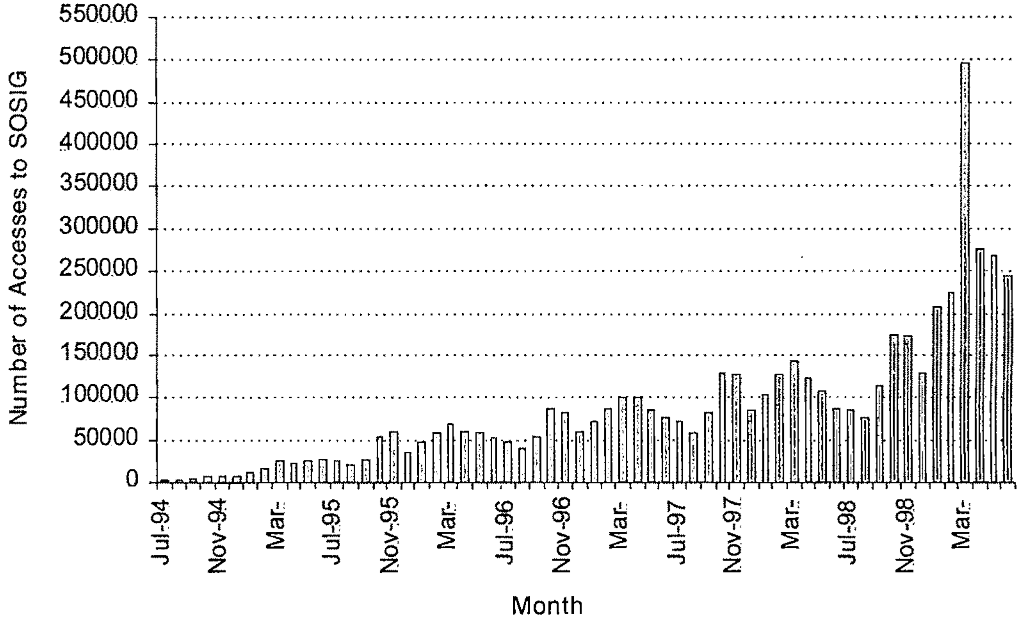
About SOSIG
SOSIG Contacts
SOSIG Training
International Collaboration
Other SOSIG Services

the 'old' SOSIG design. Further information concerning SOSIG services can be obtained from its website: http://www.sosig.ac.uk/about_us/.

Since its inception there has been a rising demand for SOSIG's services – see Figure 2. It should be noted that the data presented in the figure excludes requests from people who work on SOSIG and the downloading of images, so as to provide an indication of 'real use'. Although there has been considerable growth this is variable. Thus for the period 1994–1999, October, November and March are plainly the peak months. These peaks coincide with the beginnings and ends of the university spring and autumn terms. This trend is supported by the 1999 figure for March, which shows a peak month with the highest number of accesses per month to date – 496,990.

Notwithstanding the above quantitative testament to the success of the service, users remain a valuable source of information and method of evaluating an information system. Thus a 1994 report [16] on the progress and future developments of SOSIG mentioned that 'SOSIG designers are willing to encourage the development of resource discovery tools by facilitating a dialogue between developers and users'. To this end in 1999 a small-scale survey was conducted with members of the SOSIG user group. It took place at the London School of Economics (LSE) where the group were asked to go through a set of pre-defined exercises and note down any difficulties they experienced, their general comments and anything unexpected in the way the service worked. The user group consisted of ten librarians. The exercises involved group members: browsing the SOSIG catalogue; exploring the SOSIG subject guides; finding resources using a single word search; using the advanced search options; exploring the University Social Science Department's

Figure 2: *Number of accesses to SOSIG from July 1994 to June 1999*



database, a separate database which holds records of the websites of mainly UK university social science departments. While carrying out the exercises users were invited to specify what they liked and disliked from each service provided by SOSIG. The results are summarised in Table 1.

Research methods

An online questionnaire was the main data gathering medium. Given the geographical scatter and large size of the potential user population it was the logical method. The questionnaire was made available on the SOSIG website for a period of one month, from 15 September until 15 October 1999. Significantly this period coincided with the period of peak use. The questionnaire was divided into two sections; the first section titled 'Background Information on Internet Access' was intended to obtain information on the users' background and their familiarity with the Internet. The second part titled 'SOSIG Experience' was meant to get information on how they used the SOSIG service and what they thought of it.

SOSIG survey

Characteristics of sample population

One hundred and thirty one SOSIG users responded to the survey, of whom 57% were female, 41% of the male and 2% of the respondents did not answer this question. SOSIG is used by all age ranges covered by the questionnaire (17–65+), although it is especially popular with the 25–34 age group and not so popular with those over the age of 54 (Figure 3). Regarding occupation, the largest group of respondents proved to be information scientists, accounting for 30% of the sample. Proving yet again, that despite their relatively low numbers in the population as a whole, information scientists frequently turn out to be a system's (even an end-user one) heaviest users. Of the rest of the respondents 16% were undergraduate, postgraduate, or research students, 12% were lecturers or professors, and 11% were research staff, such as research fellows or assistants (Figure 4). The category 'others' includes occupations such as webmasters and managers. The 'don't know' and blank categories represent the

Table 1: *SOSIG user group evaluation*

Exercises	**Liked**	**Disliked**
Browse	• Flexibility • Easy to use • The fact that someone can search different sections by resource type or alphabetically	
Explore the SOSIG subject guides	• Nicely laid out • Easy to use • Ability to conduct more precise and specific searches	• Subject terms seem to be objective
Find resources using a single word search	• Easy to use • The fact that keywords are highlighted in the resource description	• No sense in how the 'hits' records are ordered
Use the advanced search options	• Easy to use • Ability to conduct more precise and specific searches • Ability to conduct more sophisticated searches	• No sense in how the 'hits' records are ordered
Explore the University Social Science Department's database	• Easy to use • Ability to have all Social Sciences Departments listed in one place	• Not clear what exactly the University Social Science Department's database consists of
Use of SOSIG thesaurus	• Ability to retrieve narrower, broader, and related terms • Easy to use	• The fact that alternative language terms are not included

respondents who either did not give an answer or for whom the responses were incomplete.

Obtaining information from the Internet

Academics are seen generally to be very heavy Internet users. The vast majority of respondents said they were regular Internet users: 86% said they used the Internet daily and 8% weekly. Only 6% used the Internet occasionally. Both men and women proved to be regular Internet users: 93% of women used the Internet on a daily or weekly basis while the figure for men was 91%.

Despite high volumes of use (or because of it), two-thirds of users said they experienced problems when obtaining information from the Internet. These problems varied from the quality of information on the Net and the difficulty of users in obtaining information that met their information needs, to technical issues such as how long it takes for a page to be downloaded. Men and women users both experienced problems with the Internet, *but* women seemed to have the greater difficulties: 71% of women indicated that obtaining information from the Internet was either difficult or moderately easy while the equivalent

Figure 3: *Age of SOSIG users*

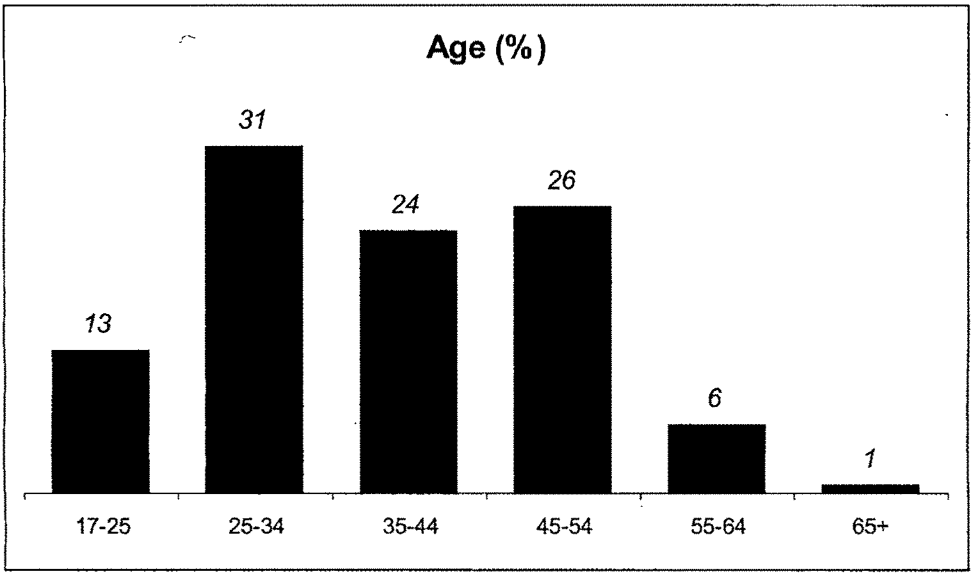


Figure 4: *Occupation of SOSIG respondents*

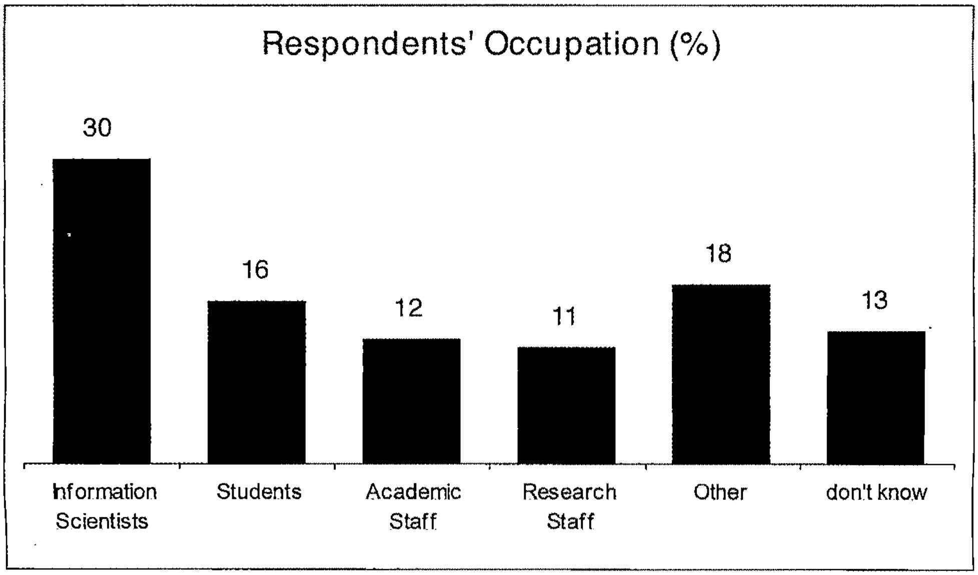


figure for men was 54% (Figure 5). When users were asked to specify their problems (see Figure 6) a sizeable majority (64%) said the main problem was overload (too much information available); 45% mentioned that they did not have the time required to search for information they needed, and a similar proportion (44%) cited slow response times. Both men and women thought overload to be the main problem but it concerned women

Figure 5: *Obtaining information from the Internet – women and men*

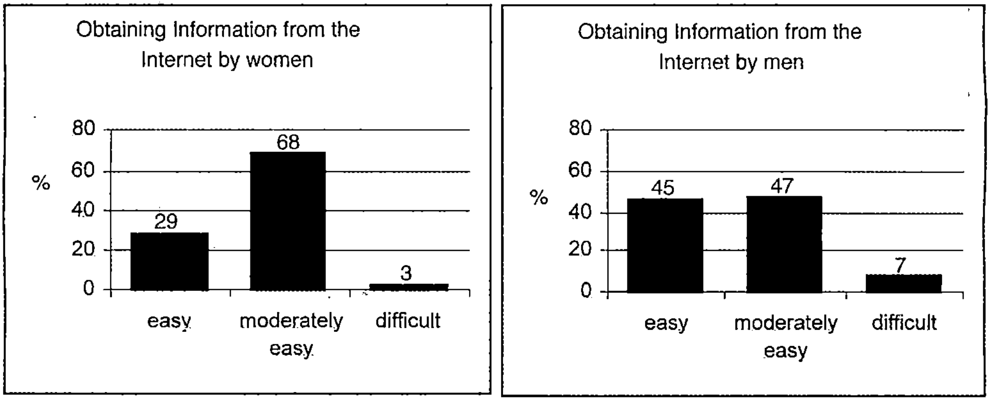
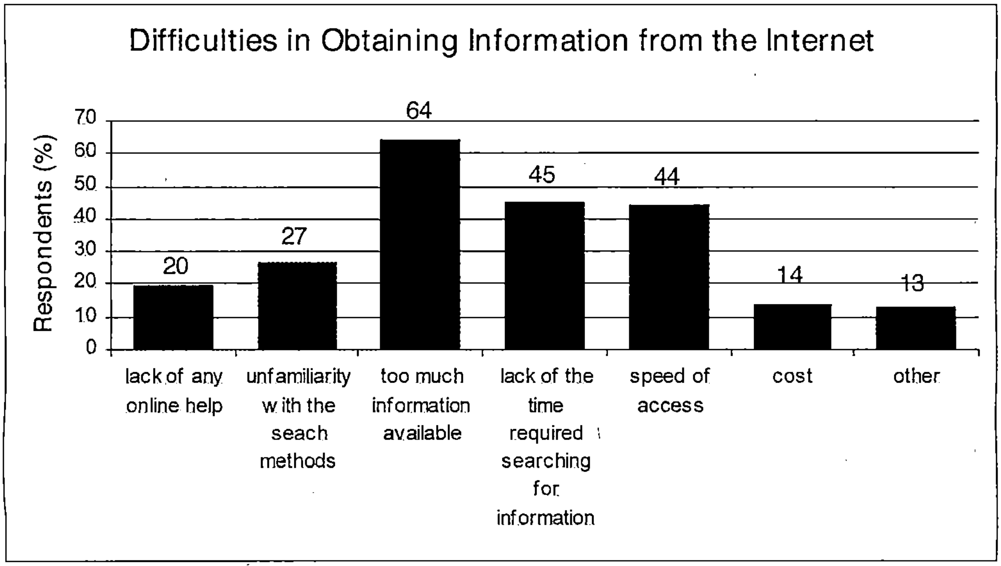


Figure 6: *Difficulties in obtaining information from the Internet*



rather more than men (Figures 7 and 8). Men appeared to be more concerned with speed than women.

Frequency of SOSIG use

A good deal of SOSIG use – unlike general Internet use – is irregular and light. Forty six per cent of the respondents indicated that they accessed SOSIG occasionally. Just over a third used it on a weekly basis and probably unsurprisingly, only 3% of the respondents used SOSIG every day (Figure 9). Interestingly, those aged between the age of 25 and 34 accounted for three-quarters of respondents who used SOSIG daily. Men also used the service more frequently: 80% of them used the service on a daily, weekly or monthly basis, whereas only 33% of females used it frequently. Regarding respondents' occupation information scientists were the most regular users with 79% of them using SOSIG on a daily, weekly, or monthly basis (Figure 10).

Figure 7: *Difficulties experienced by women in obtaining information from the Internet*

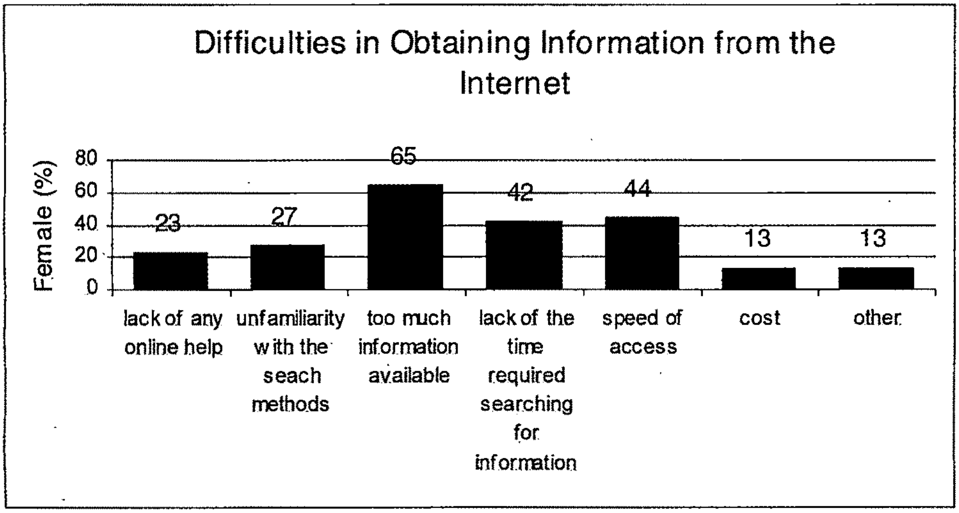
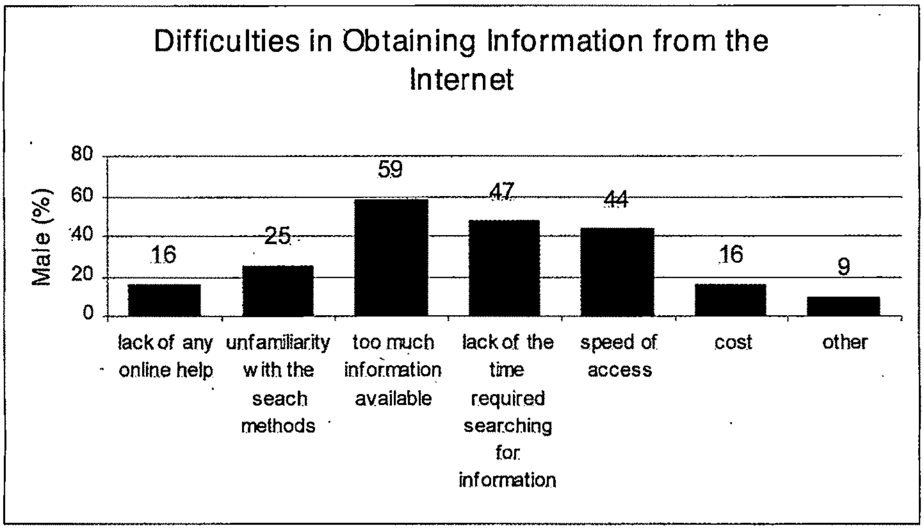


Figure 8: *Difficulties experienced by men in obtaining information from the Internet*



Reasons for using SOSIG

People may use SOSIG for a wide variety of purposes and they were asked why they used SOSIG. The options they were given were: research – identifying resources for writing a paper or carrying out a study; supporting teaching – identifying resources for supporting a lecture; or personal use – identifying resources for keeping up-to-date without a specific purpose. There was also the 'other' option where respondents could indicate any other reason for using SOSIG. Of the respondents 67% said that research was the main reason for using the service (Figure 11). Rather more men (68%) used SOSIG for research than women (64%).

Figure 9: *Frequency of SOSIG use*

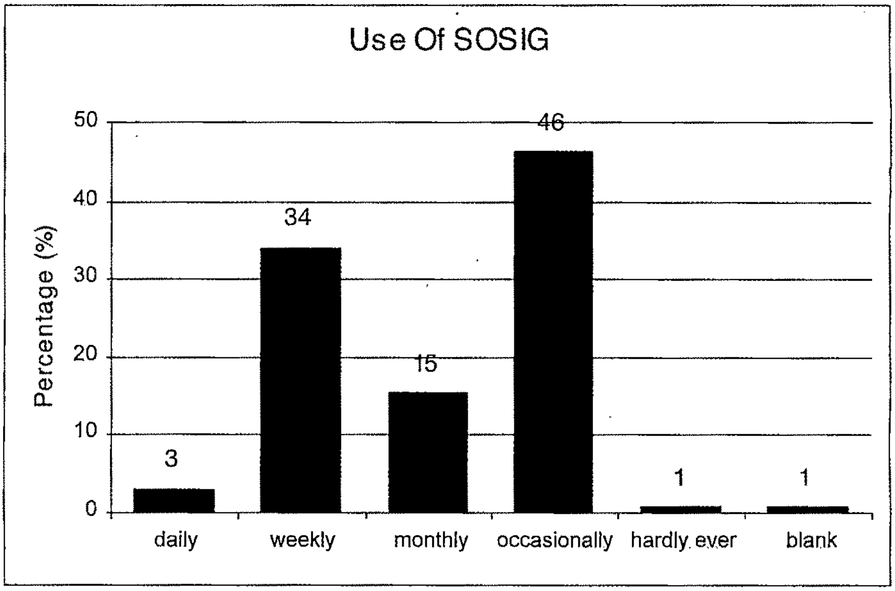
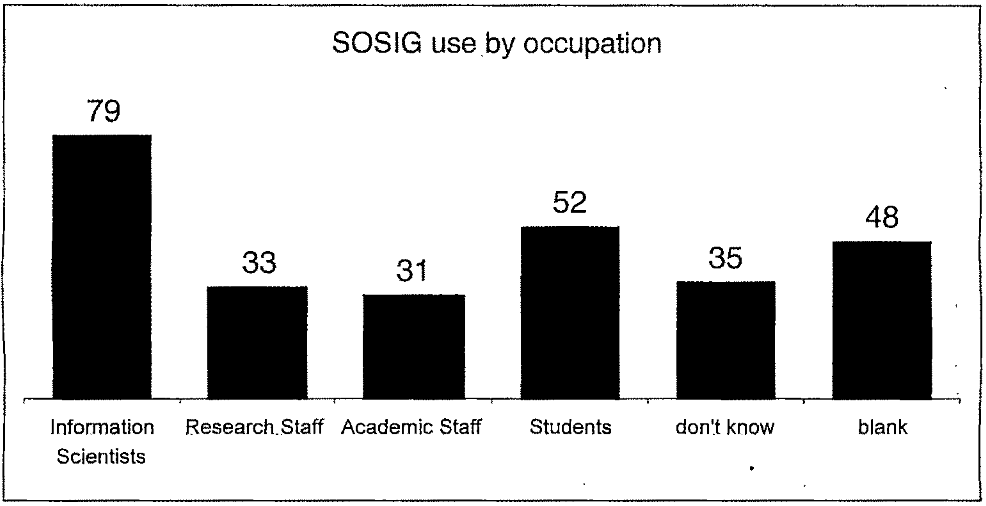


Figure 10: *SOSIG use by occupation of user*

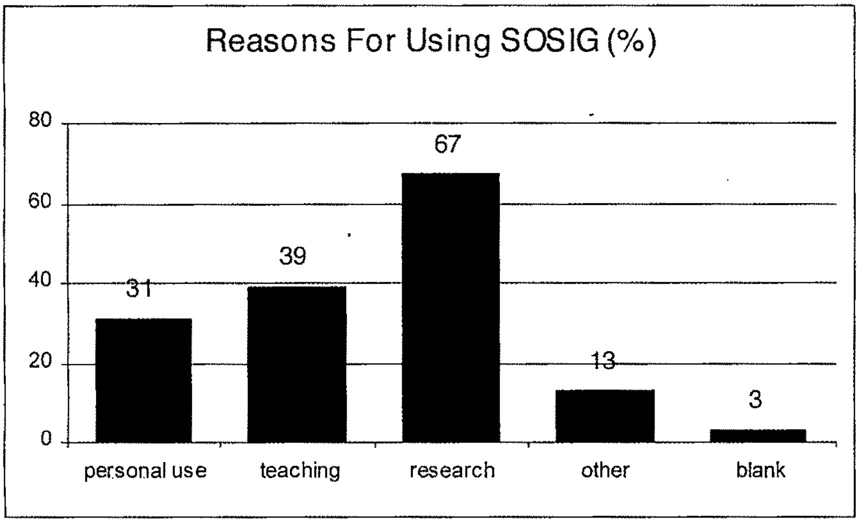


Evaluation of specific SOSIG services

The SOSIG service provides its users with a number of services including online help, browse and search facilities (see Figure 1). Users were asked to express their opinion on which search method they preferred and whether they had used online help or not.

Searching the catalogue is a valuable retrieval method when users know what they are looking for – a specific resource or information on a specific subject area of social sciences. Browsing offers users the chance to find all the available resources under a particular subject area, or relating to a particular geo-

Figure 11: *Reasons for using SOSIG*



graphical region. People who used the online help service were also invited to comment on it.

Search or browse

The most popular search method was direct searching – 50% of respondents preferred this method (Figure 12). If we discount the blank responses this figure rises to 59%. Browsing was preferred by 33% of respondents: 39% if we exclude the blanks. Women showed a preference for searching – 53% of them said they preferred searching as compared to 47% of men. Respondents belonging to the age groups: 17–24, 35–44, 45–54, and 65+ showed a preference for searching, while 25–34 and 55–64 aged groups preferred browsing. Of the age bands, 35–44 showed the strongest preference for searching (69% opted for this method). In contrast those aged between 55–64 were the biggest browsers – 50% said they did so. There are some big differences here that requires further investigation.

As can be seen below users' comments regarding their preference for the searching method indicated that it saved their time and provided them with more accurate and direct information.

> Searching is a more powerful method than browsing because it gives more quick and specific information.
>
> I am usually looking for something specific so I tend to search for it exclusively.

In contrast, they said that browsing allowed them to do the equivalent of a 'shelf search' and to identify resources in a specific area.

> You can find things you didn't know were there to search for because sometimes it's not clear what discipline a subject falls under.
>
> The contents are rich and one can choose according to one's needs and interests.
>
> I might see other relevant resources in passing; I won't miss any resources in a particular area.

To obtain more details on the searching method employed users were asked to specify which SOSIG search or browse options they used. Users are able to search or browse by using a variety of options: by title, by keywords, by descriptions, by country, by resource type, and by language. There is a very good spread of use of these search options. The option most frequently used

Figure 12: *Search or browse*

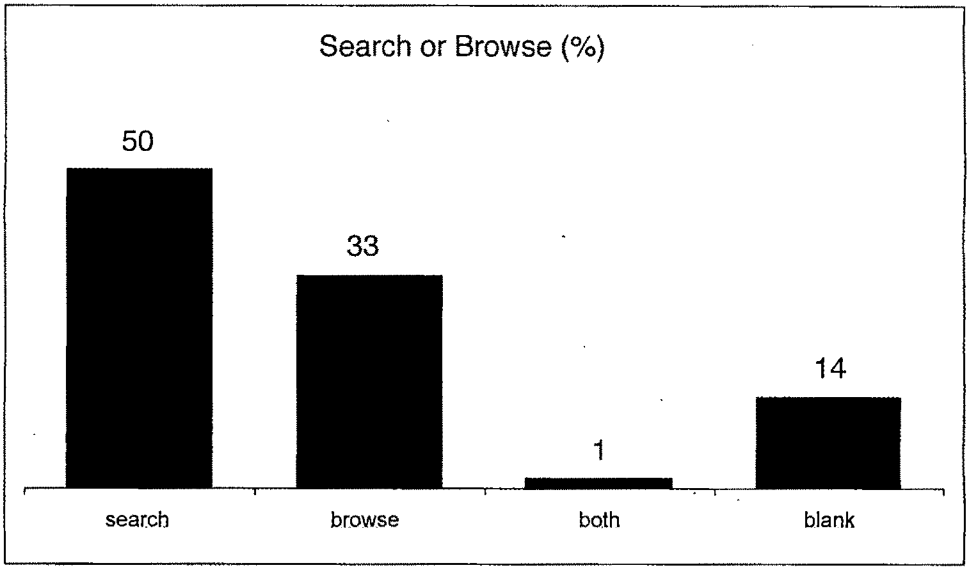


Figure 13: *Search or browse options use*

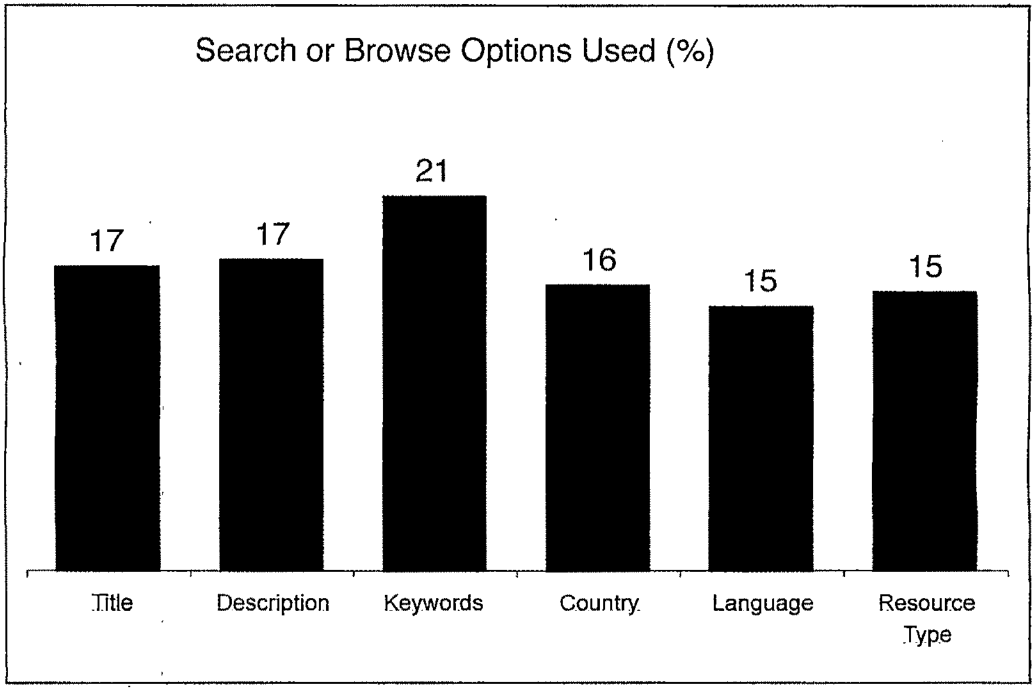


Figure 14: *Search or browse options by occupation*

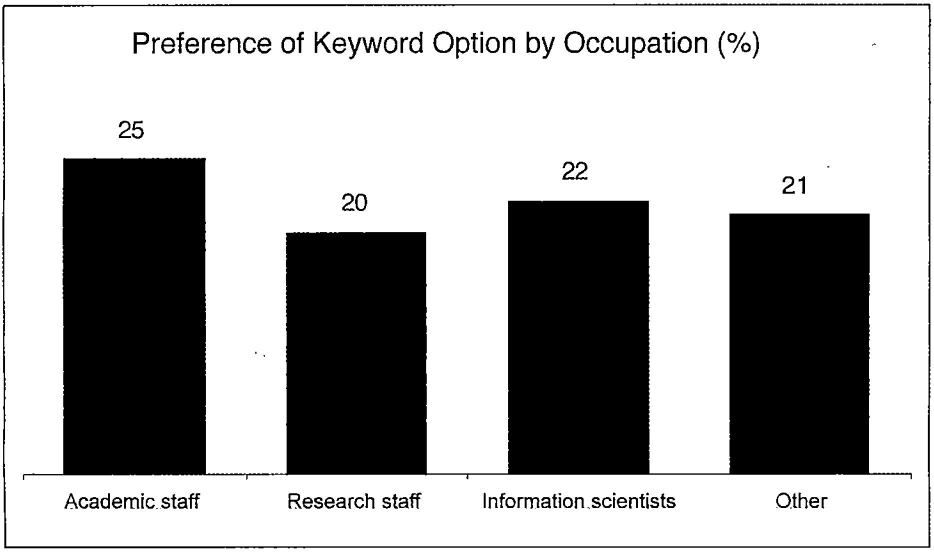
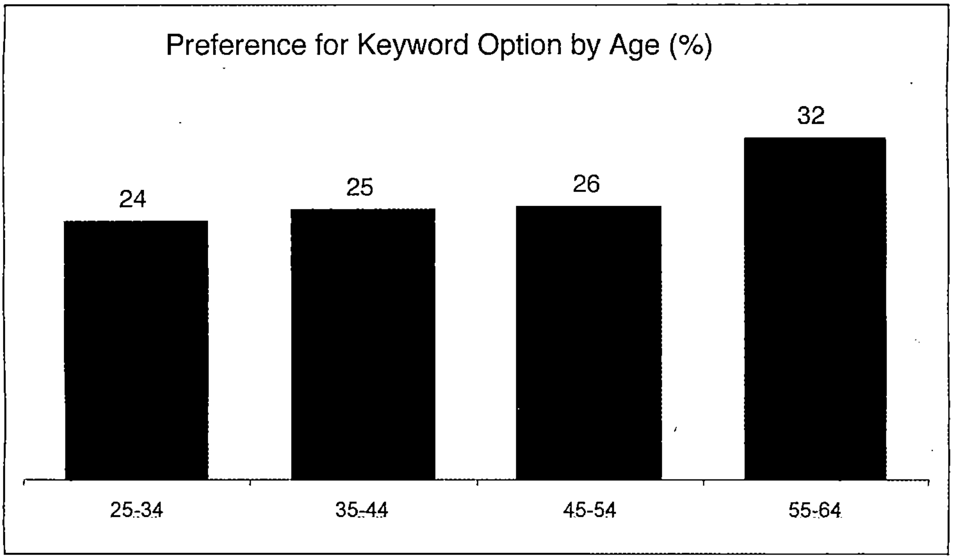
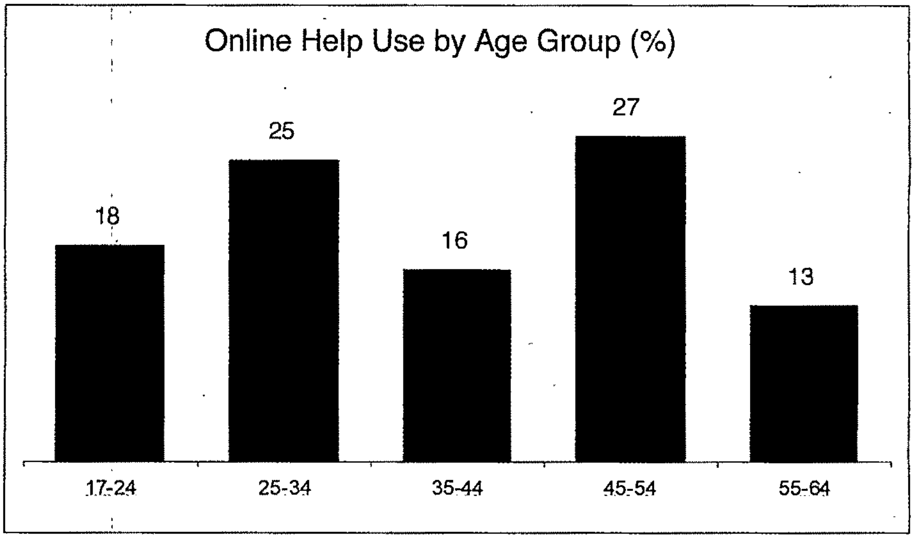


Figure 15: *Preference for keyword option by age*



was 'keywords', 21% of respondents mentioned it (Figure 13). Academic staff, research staff, information scientists, and the 'other' occupational categories preferred the 'keywords' option (Figure 14), while students used the 'keywords' and 'descriptions' options equally. Of the age groups those aged between 55–64 said the 'keywords' option was their most favourite option of obtaining information (Figure 15). Users aged between 17–24 indicated that they used the 'title', 'descriptions' and 'keywords' options equally.

Figure 16: *Online help use by age groups*



Online help
Around one in five (21%) of the respondents had called on online help whilst searching the system. Women appeared to need more help than men – 67% of women used help, but only 33% of men said they did so. The online help function seemed to have been used by all occupation groups, but mostly by information scientists: 57% of the information scientists used the online help. Online help was also used by all age groups, except for respondents aged over 65 who did not use it at all (Figure 16). Respondents were also invited to describe this service and two-thirds (64%) of the online help users stated that it is a useful facility. Notwithstanding the low number of users having used the online help, it does not necessarily imply that users do not need help during their searching or that users would evaluate the online help as a redundant service.

Discussion and conclusions

Since SOSIG was first introduced there has been a marked increase in the number of user requests. The rapid expansion of networked information and the inability of many users to evaluate this information according to their own interests account for this growth. SOSIG results revealed that women seemed to have more difficulties with the Internet than men. The fact that too much networked information is available was the most frequent problem that both women and men faced when searching the Internet. It would be premature to say whether the increase in SOSIG use will continue or not, but the general concept of an organised collection of information based on the Internet seems, for many, to be a solution to the problem of finding information on the Internet. Men and those people aged 25–34 appeared to be the heaviest users of SOSIG. However, only 3% of the respondents used SOSIG on a daily basis, while 46% of them used it occasionally. The infrequent use of SOSIG is supported by other published studies of digital resources. The primary reason given is that information seekers are unwilling to attempt new literature searching practices until they are totally convinced of their efficacy. Another reason is that the awareness of 'new' online information systems such as SOSIG is generally poor, with people usually sticking with the services that they feel most comfortable with. People are not aware of the developments that are taking place and it is still the case that little information currently exists in electronic form to justify the time and effort required to learn the skills to use them.

SOSIG supports its users by identifying authoritative and relevant information resources, varying from digitised books to education software in order that they can get appropriate information to support their research and teaching. Direct searching is the most favourable method of collecting information. This is presumably because we are largely talking about an information-knowledgeable group of users. The ubiquitous 'keyword' search proved to be the most popular. Few respondents used the online help function, although women and information scientists found it particularly valuable. This appears to support another finding that women admitted to having more problems on the Internet than men. Despite the low use of the online help function it would not be right to conclude that users do not need help and are always able to carry out successfully their search by themselves. Possible explanations might lie in the lack of awareness of the existence of online help, or the lack of time for carrying out a new search based on the 'online help' suggestions. Previous studies found that users of well-used information systems, such as BIDS, were not always using their full potential. Therefore, the availability of training to users that provided guidance on how an information system is used and which services it provides, would probably be of benefit to users of SOSIG.

References

1. Colley, A. *Computer assisted journalism: the quality issues. SOSIG: an Internet service for the 21st century launch and workshop programme, 25 February 2000.* London: National Institute for Social Work, 2000.

2. Subject Gateways. http://www.desire.org/html/subjectgateways/subjectgateways.html (visited 23 March 2000).

3. Wierenga, K. *Comparative evaluation of the subject approach to providing access to network resources.* http://www.ukoln.ac.uk/services/elib/papers/tavistock/subject-gateway-access/ (visited 18 October 1999).

4. Dix, A., Finlay, J., Abowd, G. and Beale, R. *Human – computer interaction.* New York: Prentice-Hall, 1995.

5. *Internet Library of Early Journals (ILEJ): final report.* March 1999. http://www.rsl.ox.ac.uk/ilej/papers/fr1999/ (visited 4 April 1999).

6. *Headline newsletter: hybrid electronic access and delivery in the library networked environment. What users want from the hybrid library?* 1999. http://www.headline.ac.uk (visited 9 April 1999).

7. Covi, L. and Kling, R. Organisational dimensions of effective digital library use: close rational and open natural systems models. *Journal of the American Society for Information Science*, 47(9), 1996, 672–689.

8. The STM Information System in the UK. *A study on behalf of the Royal Society, the British Library and the Association of Learned and Professional Society Publishers.* London: British Library Research & Innovation Centre, 1993. (British Library R&D Report No. 6123)

9. Rolinson, J., Meadows, A.J. and Smith, H. Use of information technology by biological researchers. *Journal of Information Science*, 21(2), 1995, 133–139.

10. Rolinson, J., Al-Shanbari, H. and Meadows, A.J. Information usage by biological researchers. *Journal of Information Science*, 22(1), 1996, 47–53.

11. Brunskill, K. *CASIAS Services: a critical evaluation of the functionality, costs, impact and value.* London: British Library Research & Innovation Centre, 1996. (British Library R&D Report No. 4)

12. Bishop, A. P. *Measuring access, use, and success in digital libraries.* 1998. http://www.Press.umich.edu/jep/04–02/bishop.html (visited 8 February 2000).

13. SuperJournal Project. *Summary of SuperJournal findings: readers.* http://www.mimas.ac.uk/sj/findread.htm (visited 17 February 2000).

14. Philip, G. Use of 'leading-edge' information system by academic chemists in the UK: Part I. The result of a preliminary investigation. *Journal of Information Science*, 21(3), 1995, 189–199.

15. Philip, G. Use of 'leading-edge' information system by academic chemists in the UK: Part II. Constraints and the need for usability engineering. *Journal of Information Science*, 22(2), 1996, 93–106.

16. Hiom, D. *The social science information gateway: a report. 1994.* (unpublished)